ED 019 477 08 VT 004 756

THE RELATIVE EFFECTIVENESS OF INFORMATIONAL FEEDBACK ABOUT SUPERVISORY AND STUDENT REACTIONS WITH BEGINNING AND EXPERIENCED VOCATIONAL TEACHERS. FINAL REPORT.

BY- OLIVER, WILMOT F.
RUTGERS, THE STATE UNIV., NEW BRUNSWICK, N.J.

REPORT NUMBER BR-6-8327 PUB DATE OCT 67
CONTRACT OEC-1-7-0396
EDRS PRICE MF-$0.50 HC-$3.28 80P.

DESCRIPTORS- *FEEDBACK, *VOCATIONAL EDUCATION TEACHERS, STUDENT OPINION, TEACHING QUALITY, TEACHER IMPROVEMENT, QUESTIONNAIRES, COMPARATIVE ANALYSIS, EXPERIMENTAL GROUPS, CONTROL GROUPS, *EFFECTIVE TEACHING, TEACHER CHARACTERISTICS, TEACHER EXPERIENCE, SUPERVISORS, BEGINNING TEACHERS, *TEACHER EVALUATION,

THE PURPOSE OF THIS STUDY WAS TO DETERMINE THE RELATIVE EFFECTIVENESS OF INFORMATIONAL FEEDBACK FROM SUPERVISORS, STUDENTS, AND STUDENTS AND SUPERVISORS COMBINED AS A MEANS OF IMPROVING THE TEACHER IMAGE OF BEGINNING AND EXPERIENCED TEACHERS. THE 286 VOCATIONAL TEACHERS INCLUDED IN THE STUDY WERE CATEGORIZED INTO GROUPS BASED ON YEARS OF TEACHING EXPERIENCE AND THEN RANDOMLY ASSIGNED TO ONE OF FOUR FEEDBACK CONDITIONS--(1) SUPERVISOR ONLY, (2) GRADE 10, 11, AND 13 STUDENTS ONLY, (3) SUPERVISOR AND STUDENTS COMBINED, AND (4) A CONTROL GROUP WHICH RECEIVED NO FEEDBACK. FORM 6 OF THE STUDENT OPINION-QUESTIONNAIRE WAS ADMINISTERED TO THE FEEDBACK GROUPS AT THE BEGINNING AND END OF A 12-WEEK PERIOD. THE DATA WERE ORGANIZED AND MAILED WITH INTERPRETIVE INFORMATION TO EACH TEACHER IN THE EXPERIMENTAL GROUPS. THE DATA WERE USED IN A 2X2X3 FACTORED EXPERIMENT WITH THE STATISTICAL METHOD OF UNWEIGHTED MEANS BECAUSE OF DIFFERENT SIZES OF CELLS. THE FINDINGS WERE--(1) THERE WERE SIGNIFICANT DIFFERENCES IN TEACHER EFFECTIVENESS AS OBSERVED BY STUDENTS BETWEEN THOSE GROUPS RECEIVING FEEDBACK FROM STUDENTS, EITHER ALONE OR IN COMBINATION WITH FEEDBACK FROM SUPERVISORS, AND THOSE WHO RECEIVED NO FEEDBACK, AND (2) STUDENT FEEDBACK IMPROVED TEACHER EFFECTIVENESS WHILE SUPERVISOR FEEDBACKS DID NOT, AND THE EFFECT OF COMBINED FEEDBACK DID NOT EXCEED THAT OF STUDENT FEEDBACK ALONE. STUDENT FEEDBACK DURING THE FIRST 10 YEARS OF TEACHING CAN BE USED AS AN EFFECTIVE METHOD OF IMPROVING TEACHER EFFECTIVENESS AS SEEN BY STUDENTS. A BIBLIOGRAPHY, THE QUESTIONNAIRE, AND RELATED RESEARCH MATERIALS ARE INCLUDED. (MM)

ED019477

FINAL REPORT
Project No. 6-8327
Contract No. OE-1-7-0396

THE RELATIVE EFFECTIVENESS OF INFORMATIONAL FEEDBACK ABOUT SUPERVISORY AND STUDENT REACTIONS WITH BEGINNING AND EXPERIENCED VOCATIONAL TEACHERS

October 1967

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

The Department of
Vocational - Technical Education
Graduate School of Education
Rutgers - The State University

VT004756

THE RELATIVE EFFECTIVENESS OF INFORMATIONAL FEEDBACK ABOUT SUPERVISORY AND STUDENT REACTIONS WITH BEGINNING AND EXPERIENCED VOCATIONAL TEACHERS

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Project No. 6-8327

Contract No. OE-1-7-0396

Wilmot F. Oliver

Rutgers - The State University

New Brunswick, New Jersey

October 1967

The research reported herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

ACKNOWLEDGMENTS

The writer wishes to express appreciation to the members of the graduate committee each of whom offered helpful suggestions as they were needed. The committee consisted of Dr. G. Franklin Stover, Chairman, Dr. Carl J. Schaefer, Dr. Charles Drawbaugh, and Dr. Bruce Tuckman.

Appreciation is extended to the vocational teachers in Delaware, New Jersey, Pennsylvania, and New York who provided the data needed in the study.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS ii

LIST OF TABLES vi

LIST OF FIGURES vi

ABSTRACT vii

Chapter

I. INTRODUCTION 1

Need for the Study
General Statement of the Problem
Scope and Limitations of the Study
Definition of Terms Used

II. REVIEW OF THE LITERATURE 9

Vital Literature
Effect of Age and Years of Teaching Experience Upon Teacher Effectiveness
Research on the Reliability of Student-Instructor Ratings
Summary

TABLE OF CONTENTS (Continued)

Chapter Page

III. DESIGN OF THE STUDY 22

Specific Statement of the Problem
General Design
The Sample
Sources of Informational Feedback
Years of Teaching Experience
Measurement of the Dependent Variable
Student Opinion-Questionnaire and Data Collection
Communicating Feedback
Analysis of Data

IV. PRESENTATION OF DATA 32

Differences Between the Control and Experimental Groups
Differences Due to Sources of Informational Feedback
Differences Due to Years of Teaching Experience

V. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS . . 43

Implications
Conclusions
Recommendations for Further Study

BIBLIOGRAPHY 50

APPENDIXES

APPENDIX A - Criteria for Selection of Opinion-Questionnaire 54

APPENDIX B - Student-Opinion Questionnaire . . . 56

TABLE OF CONTENTS (Continued)

Page

APPENDIX C - Summary of Comments by Students . . 58

APPENDIX D - Student Reaction Report 59

APPENDIX E - Personal Data 60

APPENDIX F - Letter of Explanation to Accompany Feedback 61

APPENDIX G - Instructions for Administering the Student Opinion-Questionnaire . . 62

APPENDIX H - Letter Requesting Respondents to Administer Student-Opinion Questionnaire 63

APPENDIX I - Letter Requesting Supervisor to Fill Out Opinion-Questionnaire . . 64

APPENDIX J - Confidence Intervals 65

APPENDIX K - Schools Participating in the Study . 69

APPENDIX L - Distribution of Teachers' Ages . . . 70

APPENDIX M - Geographical Distribution of Teachers 71

LIST OF TABLES

Table Page

1. Analysis of Variance of Items 1-5 (A) Teacher Experience, (B) Supervisor Reactions and (C) Student Reactions 33
2. Analysis of Variance on Items 6-10 (A) Teacher Experience, (B) Supervisor Reactions and (C) Student Reactions 34
3. Mean Difference Scores Calculated by Analysis of Variance for 10 Items Used to Evaluate Vocational Teachers by Three Groups of Raters for a Control and Three Experimental Groups . 36
4. Means Calculated by Analysis of Variance for 10 Items Used to Evaluate Vocational Teachers by Three Experimental Groups 38
5. Means Calculated by Analysis of Variance for 10 Items Used by Vocational Teachers in Three Experience Groups 40
6. Mean Difference Scores Calculated by Analysis of Variance for 10 Items Used to Evaluate Vocational Teachers by Student Raters Using Two Time Intervals (12 Weeks and 40 Weeks) . . 41

LIST OF FIGURES

Figure

1. Assignment of Instructors to Treatment and Experience Groups 25

ABSTRACT

The purpose of this study was to determine the relative effectiveness of informational feedback from supervisors, students, and students and supervisors combined as a means of improving the teacher image of beginning and experienced teachers. The problem was to test three hypotheses:

1. that teachers exposed to informational feedback would change more than those not so exposed.

2. that teachers exposed to different sources of informational feedback would not change to the same degree, i.e. the sources have a differential effect.

3. that the amount of teacher change as a result of informational feedback would be inversely related to the amount of his teaching experience.

Instructors were categorized into three groups based upon years of teaching experience and then randomly assigned to one of four feedback conditions: (1) supervisor only, (2) students only, (3) supervisor and students combined, and (4) a control group which received no feedback. Two hundred eighty-six instructors from vocational departments in New Jersey; Wilmington, Delaware; Bucks County, Pennsylvania; and Rockland County, New York were utilized in the experimental study. The pretest was administered during the fall semester of 1965 and the posttest was administered 12 weeks later. Reactions about teacher behavior were gathered from supervisors and also students enrolled in grades 10, 11, and 13. Upon receipt of the answered questionnaires, the data were organized into a report and mailed with interpretive information to each teacher in the experimental groups.

The data gathered were used in a 2 X 2 X 3 factorial experiment. The statistical method of unweighted means was used to analyze data because of disproportionality in cell strength.

The findings of the study were as follows:

1. There were significant differences in teacher effectiveness as observed by students between those groups receiving feedback from students, either alone or in combination with feedback from supervisors, and those who received no feedback.

2. The various sources of informational feedback were not equally effective. Student feedback improved teacher effectiveness while supervisor feedback did not. The effect of the combined feedback did not exceed that of student feedback alone.

3. The most experienced teachers (11 years and over) were least receptive to feedback as compared to the intermediate experienced group of teachers (4-10 years) and those teachers with limited teaching experience (1-3 years). The effect of feedback on the intermediate and less experienced teachers was approximately equal but greater at the .05 level of significance than the most experienced group.

Based upon these findings, student feedback during the first 10 years of teaching can be used as an effective method of improving teacher effectiveness as seen by students.

CHAPTER I

INTRODUCTION

Appraisal of teaching has existed in some form for thousands of years. Wherever students, fellow teachers, administrators and parents gather, teachers are rated. In the informal situation these "ratings" take on the form of rumor and gossip and probably have had a far greater effect upon the reputation of teachers than has been commonly realized. Even though evaluation over the years has become somewhat formalized, the evaluation of teaching has not reached a high level of sophistication when compared with the controlled laboratory of the physical sciences.

Too often one of the most important purposes of evaluation has been forgotten--to help the evaluated identify his strengths and weaknesses, and to help him understand and accept himself in terms of abilities, patterns of interest, background preparation, and emotional make-up. As a result of having such information, it is believed that teachers can and will set more realistic goals for themselves.

In day-to-day teaching, the instructor receives feedback by: (1) observing his class, (2) talking informally with his pupils, (3) speaking with other teachers, (4) having conferences with parents, (5) talking with interested people in the community, (6) consulting his principal or supervisor, (7) as well as by giving tests. For the most part, these sources of feedback serve fairly well for the majority of teachers. However, it is reasonable to assume that there is room for improvement. One rather obvious way to bring about such improvement is to manipulate conditions so as to improve the teacher's accuracy in perceiving pupils' perceptions of his teaching.

The value of informational feedback to a teacher from his pupils might be expressed with an analogy: An artillery battalion firing on an enemy target will not improve its accuracy without the crucial information relayed back by a forward observer. As soon as the gunners receive information concerning the range and azimuth, the accuracy improves with each correction until finally the target is pin-pointed. The improvement is due to knowledge of results or "feedback."

In everyday teaching, the teacher may be thought of as firing his behaviors, gestures, facts, and concepts at his pupils. How much they understand, are motivated, and learn may well depend in part at least, on the kind and amount of feedback received from pupils in day-to-day interaction with them.

McCall concluded in a study of teacher merit:

> At last we find some professional competent judges of teaching skill, namely teachers' pupils, especially after they have been taught by the teacher for nearly a year. Out of the mouths of children comes more accurate judgment of teachers than that rendered by their peers and supervisors, and, if our criterion is valid, they appear to have a truer idea of what constitutes good teaching than professors of education.[1]

Need for the Study

It is not possible for teacher education institutions to produce completely finished teachers. Furthermore, improvement in teaching does not necessarily follow from experience alone. There has always been a need to help teachers become competent on the job. Supervisors, workshops, conferences, and various in-service teacher education courses are used to meet this need in part. As the shortage of vocational teachers continues, even greater efforts will be needed to aid teachers in fundamental orientation to their task. If such aid is not forthcoming, the teacher may fail to overcome the rather obvious limitations of his pre-service and in-service preparation. There is some evidence to support the belief that increased awareness by the teacher of his pupils' or supervisor's perceptions of his actions can result in instructor improvement.[1]

---

[1] W. A. McCall, Measurement of Teacher Merit, Bulletin No. 284 (Raleigh, North Carolina: State Department of Instruction, 1959), 29 pp.

Innumerable abilities, attitudes, motives, and psychological traits are believed to relate to the competence of teachers. These characteristics are presumed to characterize the teacher in a consistent manner over a period of time, and are believed by many to explain his behavior in response to most situations. The nature of the characteristics are such as to make direct observations impossible in the same way that behavior can be observed. However, one can observe the putative effects of such qualities in teacher behavior. One can ask supervisors, fellow teachers, and others to report or rate these qualities. At the present time there is no ideal method for measuring any of these teacher-properties. Even such a well established property as intelligence has shown a discouragingly low correlation with teacher effectiveness. This has been equally true of almost every teacher property investigated with respect to a meaningful relationship with pupil change.

Many studies have already been done to discover factors that make a difference in the effectiveness of teaching. In much of the research the characteristics of teachers (abilities, attitudes, motives, and psychological traits, etc.) have served as independent variables at Time 1, while the achievement of pupils served as the dependent variable at Time 2. These studies have not yielded many positive findings and the relationships obtained have generally been low and inconsistent from one study to the next.

---

[1] Roy C. Bryan, Reactions to Teachers by Students, Parents, and Administrators, United States Office of Education, Cooperative Research Project No. 668 (Kalamazoo, Michigan: Western Michigan University, 1963), p. 43.

The reason for this lack of consistency may be that the characteristics of the teachers measured in these studies (such as attitudes, intelligence, or perceptual accuracies) still did not vary enough to make much difference in the kind of dependent variable with which the researchers have been concerned.

A second explanation is that the studies concentrating solely on prior characteristics of the teacher fail to take into account a very important factor in the teaching-learning process, namely, the characteristics of the pupils. Their abilities, interests, needs, values, and perceptions have been overlooked as possible determiners of classroom phenomena. Since teaching is generally viewed as an interactive process, classroom events cannot be solely accounted for in terms of the characteristics and behaviors of teachers. Therefore, it may be seen that those events must be considered to be outcomes in which pupil variables have an important effect. Any adequate design for research in the area of informational feedback and learning phenomena must of necessity include provisions for the characteristics, behaviors, and perceptions of the pupils in relation to the teacher.

General Statement of the Problem

The general area of investigation in this research was to study the relative effectiveness of informational feedback from different sources to vocational instructors. Three sources of feedback were compared: students only, supervisors only, and a combination of both supervisors and students. An additional aim was to determine whether years of teaching experience made a significant difference in the results of informational feedback from the different sources. Effectiveness of feedback was defined as changes in teaching behavior, inferred from changes in the students' ratings from pretest to posttest.

The first hypothesis tested in this study involved feedback from students and is essentially a replication of studies done by Roy C. Bryan[1] and others. Only the interval of feedback differs significantly from Bryan's study. It was much shorter than that used in Bryan's study which covered a full two-year period with two posttests. This research was an effort to compare the effectiveness of supervisory feedback with student feedback in order to bring a new dimension to the study.

In this investigation there were two critical controls:

A. A control group which was administered a pretest and posttest with no feedback of information.

B. A posttest only group.

In testing the first hypothesis, student feedback, supervisory feedback, and combined supervisory and student feedback were compared with those receiving no feedback in the control group.

In the second hypothesis the three sources of feedback were compared to each other to determine significant differences.

Finally, in the third hypothesis three experience groups consisting of beginning teachers (1-3 years), experienced teachers (4-10 years), and experienced teachers (11 years and over) were compared to determine whether or not significant differences were present due to years of teaching experience.

Previous studies have tended to indicate that teachers are indeed sensitive to informational feedback from students. A review of the literature also shows that students constitute a pool of reliable observers who are in a favorable position to observe changes in the behavior of their teachers, since they are present at all times.

---

[1] Roy C. Bryan, Reactions to Teachers by Students, Parents, and Administrators, United States Office of Education, Cooperative Research Project No. 668 (Kalamazoo, Michigan: Western Michigan University, 1963).

This present study did not attempt to use student achievement as a criterion. Change in pretest to posttest scores on a student opinion questionnaire were used. Although two of the experimental groups received supervisory reactions, student ratings were used as the measuring instrument. The studies reviewed indicate that pooled student reactions tend to be more sensitive to a change in teacher behavior than a single supervisory reaction. Hermann Remmers[1] concluded that the reliability of rating of teachers by students is a function of the number of raters in accordance with the Spearman-Brown prophecy formula. His studies indicated that if 25 or more student ratings were averaged, they are as reliable as the better educational and mental tests available at present. By using student ratings as the measuring instrument under all conditions, there was no tendency to confound the results due to differences in sensitivity between student and supervisory ratings.

For the purpose of this study, reaction to informational feedback was defined as the difference between pretest scores and posttest scores on the student opinion questionnaire. It was demonstrated by comparison that teacher change at the time of posttest was a function of the informational feedback he received.

Scope and Limitations of the Study

The study was limited to include only those vocational teachers from public secondary schools and technical institutions in New Jersey; Wilmington, Delaware; Rockland County, New York; and Bucks County, Pennsylvania.

---

[1] Hermann H. Remmers, "Rating Methods in Research on Teaching," Handbook of Research on Teaching, ed. N. L. Gage (Chicago: Rand McNally and Co., 1963), pp. 329-73.

The sample included vocational teachers in federally reimbursed programs in the fields of trade and industrial education, technical education, vocational agriculture, and distributive education.

The sample included those students enrolled in grades 10, 11, and 13. Grade 9 was not used because many schools started vocational training at the 10th grade level.

The study was conducted during the period October 20, 1965 to October 20, 1966.

The investigator relied upon ratings by students and supervisors. The student opinion questionnaire used in this study was published by Western Michigan University. Permission was granted to use this instrument in the present study. A copy of the questionnaire is shown in Appendix B.

Definition of Terms Used

Feedback - A summary report of student and supervisor reactions used to inform the teacher concerning the opinions of his supervisor and/or students on each of 14 questions. See Appendix B.

Mean Difference Scores - The difference between posttest and pretest student judgments computed by analysis of variance using the Unweighted Means method.

Ratings - The average of the scale steps on the student and/or supervisor reaction report. See Appendix B.

Supervisor - One who is directly responsible for the teacher being observed, usually a principal, vice-principal, or director principal.

Teacher Behavior - The teachers' mode of action as reported on the 14 items on the "Student Opinion Questionnaire," by students and/or supervisors.

Vocational Teachers - Those teachers teaching trade or occupational skills to vocational classes in state and federally reimbursed programs.

CHAPTER II

REVIEW OF THE LITERATURE

The relevant literature is reviewed under four headings: (1) Summary of Related Studies, a review of the pioneer efforts in the use of student reaction reports to improve teacher effectiveness; (2) Vital Literature, a consideration of six research reports directly concerned with student-instructor ratings; (3) Effect of Age and Years of Teaching Experience upon Teacher Effectiveness, a review of selected studies about teacher resiliency as it related to age and years of teaching experience; and (4) Research on the Reliability of Student-Instructor Ratings, an evaluation of selected studies and research reports on the reliability of student-instructor ratings.

The practice of collecting and analyzing ratings of teachers from their pupils has had a moderate vogue for more than forty years. There have been many advocates of this practice and among the alleged values has been the improvement of teacher behavior. In earlier studies by such investigators as Ward, Remmers, and Schmalzried,[1] no control group was used.

Morsh and Wilder from their review of pupil ratings, concluded that:

> There appears to be considerable opinion that properly used, student rating has value in bringing about instructor improvement. For example, Schutte (1926), Clem (1930), Flinn (1932), Riley et al.

---

[1] William D. Ward, Hermann H. Remmers, and Newell T. Schmalzried, "The Training of Teacher-Personality by Means of Student Ratings," School and Society (1941), 189-93.

> (1950), and Ebel (1952), after having students rate instructors in one form or another, state generally without adequate research evidence that student rating enables instructors to evaluate their courses and teaching performances and their students' opinions provide a better basis for self-study and instructor self-improvement than do the opinions of supervisors.[1]

Vital Literature

Since 1927 Hermann H. Remmers has done a considerable amount of research on student ratings of instructors. Major generalizations from his research follow:

1. Reliability of ratings of teachers by students is a function of the number of raters, in accordance with the Spearman-Brown prophecy formula. If 25 or more student ratings are averaged they are as reliable as the better educational and mental tests available at present.

2. Grades of students have little if any relationship to their ratings of instructors who assigned the grades.

3. Alumni ten years after graduation agree very closely (rank order rho = .92) with on campus students on the relative importance of ten teacher characteristics.

4. Alumni ten years after graduation agree substantially (r's ranging from .40 to .68) with on campus students in their average ratings of the same instructors.

5. Halo effect, if present in ratings by such instruments as the Purdue Rating Scale for Instructors, is insufficient to raise the intertrait correlations to

---

[1] Joseph E. Morsh and Eleanor W. Wilder, "Identifying the Effective Instructor: A Review of the Quantitative Studies, 1900-1950," Air Force Personnel and Training Research Center, Research Bulletin, Tr-54-44 (San Antonio, Texas: Lackland A.F.B.), p. 39.

unity when corrected for unreliability of the ratings. Evidence indicates that students discriminate reliably among different aspects of the teacher's personality and of the course.

6. Little if any relationship exists between students' rating of the teacher and the difficulty of the course.

7. The sex of student raters bears little or no relationship to their ratings of teachers.

8. The cost in time and money of obtaining student rating of teachers is low; in fact, considerably lower than the cost of administering a typical standardized educational test of some comprehensiveness.

9. Popularity in extra class activities of the teacher is probably not appreciably related to student ratings of that teacher.

10. Teachers with less than five years experience tend to be rated lower than teachers with more than five years experience.

11. The sex of the teacher is in general unrelated to ratings received.

12. Students are more favorable than instructors to student ratings of instructors, but more instructors than students have noticed improvement in their teaching as a result of student ratings.[1]

In May of 1947, 6681 Brooklyn College students each rated five of their teachers. The Rutgers questionnaire was used and no appreciable differences in instructors' scores were found to exist because of size of classes, sex, course grades received by students, and whether or not the course was elective. Although the researcher of this study did not use a control group and made no attempt to adjust for initial differences in instructors, he made several significant contributions to the research on student ratings. Several findings not reported in other studies were: (1) students with low scholastic standing tended to rate their instructors more rigorously than those with a relatively higher academic average; (2) all of the students agreed that the faculty excelled on one attribute--"knowledge of subject"; (3) they indicated a relatively poor opinion of the faculty on "encouragement to thinking"; (4) with but one exception, younger instructors were rated superior to older instructors. Only on knowledge of subject did the older men excel. (5) Those holding Ph.D.'s surpassed all others in nine out of ten qualities of good teaching; (6) published research appeared to have a real bearing upon student conceptions of the good teacher.[1]

---

[1] Hermann H. Remmers, "Rating Methods in Research on Teaching," Handbook of Research on Teaching, ed. N. L. Gage (Chicago: Rand McNally and Co., 1963), pp. 329-73.

Since the preceding studies were reported, at least four studies of the effect of student ratings on teachers have been undertaken. In 1957, Marjorie Savage[2] investigated the effects of student ratings on junior high school teachers of home economics. In her study the subjects were student teachers who, in the experimental group, tabulated their own teacher's ratings and then discussed them with the supervising teacher. A factor that weakened the design in Savage's experiment was the fact that the first ratings were taken only five days after the student teacher began to teach, while the interval between first and second ratings was only twenty days. In a summary of the review of other studies, the interval between feedback and second ratings was an important variable in relation to the effect of feedback. Savage failed to exploit the advantages of analysis of variance for controlling initial differences between groups in relevant variables. In her results the trend was not in the hypothesized direction, nor was it statistically significant.

---

[1] John W. Riley, Jr., Bryce F. Ryans, and Marcia Lifshitz, The Student Looks at His Teacher (New Brunswick, N.J.: Rutgers University Press, 1950), pp. 80-107.

[2] Marjorie Savage, "Changes in Student Teachers Through the Use of Pupil Ratings" (unpublished Ed.D. dissertation, University of Illinois, 1957).

Another undertaking relevant to the proposed research in this study is that of Nathaniel L. Gage, Philip P. Runkel, and B. B. Chatterjee.[1] This is one of the few empirical studies concerning the influences of feedback on teacher behavior. The effects of feedback from pupils to teacher were studied. Comparison of experimental and control groups of sixth grade teachers indicated that when teachers were provided with information obtained from their pupils regarding how the pupils described their actual teacher and how they described their ideal teacher on 12 items of teacher behavior, (1) teachers' behavior changed--as indicated by subsequent pupil descriptions of their actual teacher--in the direction of the pupils initial descriptions of their ideal teacher, and (2) the teachers receiving feedback became increasingly accurate in predicting their pupils' description of their teacher. The results of this study indicated that there was a relationship between teacher change caused by feedback and the interval between feedback and post ratings of the teachers. The groups with the longest interval approached their pupils pre-ideal teacher most closely.

Bryan also sought an answer to the question: "To what extent can improvements in teacher effectiveness as judged by students be brought about through the use of written student reactions?"

In elaborating upon this question, Bryan stated:

> Testimony to the effect that student reactions have been helpful to individuals and groups is plentiful. Not so numerous are reports of

---

[1] Nathaniel L. Gage, Philip J. Runkel, and B. B. Chatterjee, "Equilibrium Theory and Behavior Change: An Experiment in Feedback from Pupils to Teachers," Bureau of Educational Research, College of Education (Urbana, Ill.: University of Illinois, 1960), 129 pp.

> improvements based on a study of favorable changes in average ratings over a period of time. One of these was made by Wilson, who stated, on those topics on which instructors had made a systematic effort to improve, the June averages were about 25 percentile points above those in December. . . . Starrak found that rating by students increased 'quite materially' with each successive rating over a two year period.[1]

Bryan's research was a longitudinal study. In the spring of one year he elicited the student reactions from the classrooms of more than 75 teachers classified as the "experimental" group. In this research, Bryan finished his study with a great imbalance between the experimental and control groups with respect to the number of years of teaching experience represented by teachers in these groups. The experimental group had twice as many teachers in the 1-5 years of experience group, while the control group had more than twice as many in the 21 year and up group. It is the opinion of this investigator that years of teaching experience is an important variable and does affect the results. The data presented by Bryan in this study very clearly indicate that the feedback of information about student reactions can be used as a means of improving effectiveness as seen by students. Other significant conclusions from Bryan's report are:

1. The image of a teacher held by students usually has much in common with the image held by administrators and parents.

2. The image one group of students has of a teacher is usually very similar to that held by other groups of students.

3. Even though no significant correlation was found between ratings of high school students (or administrators)

---

[1] Roy C. Bryan, Reactions to Teachers by Students, Parents, and Administrators, Cooperative Research Project No. 668 (Kalamazoo, Mich.: Western Michigan University, 1963), p. 58.

and student gains in the form of subject matter learned, high correlation was found between teacher prestige with students and development of interest in and liking for the subject of chemistry.

4. A large percentage of high school teachers can use information gained from student reaction feedback as a means of improving their image with students.

5. A teacher's best chance of gaining an improved image with students rests not in waiting for them to mature, but rather in increasing his prestige with students currently in his classroom.[1]

Effect of Age and Years of Teaching Experience Upon Teacher Effectiveness

The study of "Characteristics of Teachers" by Ryans[2] indicated that there is little doubt concerning the difference between teachers of different age groups with respect to a number of characteristics. Some of the results of the teacher characteristics study indicated significant differences in patterns of teacher behavior for teachers in different age groups. Generally speaking, teachers fifty-five years of age and above were at a distinct disadvantage when compared to younger teachers--except from the standpoint of systematic and businesslike behavior and learning--centered traditional education viewpoints.

---

[1] Roy C. Bryan, Reactions to Teachers by Students, Parents, and Administrators, Cooperative Research Project No. 668 (Kalamazoo, Mich.: Western Michigan University, 1963), p. 58.

[2] David G. Ryans, "Characteristics of Teachers," Contemporary Research on Teacher Effectiveness, ed. Bruce J. Biddle and William J. Ellena (New York: Holt, Rinehart and Winston, Inc., 1964), pp. 81-82.

A study by Peterson[1] is also relevant to the question of the effects of age upon teaching effectiveness. He found that teachers seemed to recognize a short orientation period as a "best teaching period," while still young and vigorous. He found that a decline set in at the age of thirty-five to forty. Beginning in their thirties the teachers in the study, almost without exception, expressed concern about losing the intimate informal contact which they had previously enjoyed with their students. In comparison, middle-aged teachers accepted this increased distance and had established a parent-like authority role. Older teachers, however, tended to complain about "students--getting worse each year." This study indicated that a critical point in the teachers career seems to occur in the 30-40 year age category. This involves intensification of professional interests and loss of intimacy with the student. Previous studies of teacher role have shown the teacher as an ageless abstraction. Peterson contended that such views ignore the important facts of aging, commitment, and job satisfaction.

The results of an investigation of "Teaching as Problem-Solving Behavior" by R. L. Turner, provided considerable evidence that:

> The very early years of teaching experience provide the greatest rise in teaching-task performance as evidenced by differences in

---

[1] Warren A. Peterson, "Age, Teachers Role, and the Institutional Setting," Contemporary Research on Teacher Effectiveness, ed. Bruce J. Middle and William J. Ellena (New York: Holt, Rinehart and Winston, Inc., 1964), p. 311.

> performance between fully prepared but inexperienced teachers and teachers with no more than three years of experience. There was little evidence to suggest that performance changed greatly for the average teacher, after the third year.[1]

Research on the Reliability of Student-Instructor Ratings

Student ratings of instructors appeared on the scene about 1923. There has been a steady increase in the rise of such ratings since that time. Many dissatisfactions concerning the reliability of student ratings have been voiced. This relates to the ability of the student to make unbiased judgments concerning a teacher's performance. Since this is a crucial objection regardless of the groups upon which it is based, an attempt will be made to document studies which have shown that in the main, these objections have not been well founded in fact.

Grades for example have shown little if any relationship to students' ratings. Starrak[2] administered his scale to over forty thousand subjects and found that the grades a college student gets does not appreciably affect the rating given by him to his instructors. A more recent study by John W. Riley[3] et al. at Brooklyn College in 1949, did show the students with low scholastic standing to be somewhat more critical in their judgments than the better scholars. Although there was no bias concerning the teachers' competence, there was a definite bias with regards to fairness on examinations. In general, this bias was not great, amounting to five or six points on a one hundred point scale. Remmers[1] also found that the relation between students' grades and their attitudes toward instructors to be a negligible (.070). Hudelson[2] found correlation coefficient of .19 for students' ratings of college instructors and student grades by their instructors.

---

[1] R. L. Turner, "Teaching as Problem Solving Behavior," Contemporary Research on Teacher Effectiveness, ed. Bruce J. Biddle and William J. Ellena (New York: Holt, Rinehart and Winston, Inc., 1964), p. 124.

[2] J. A. Starrak, "Student Rating of Instruction," Journal of Higher Education, V (February, 1934), 88-90.

[3] John W. Riley, Jr., Bryce F. Ryan, and Marcia Lifshitz, The Student Looks at His Teacher (New Brunswick, N.J.: Rutgers University Press, 1950), p. 85.

In 1935, Heilman and Armentrout[3] had over two thousand students rate forty-six college teachers on the Purdue Rating Scale. The reliability of these ratings by college students was approximately .75. They also reported that, "factors of class size, severity of grading, the student's interest in the course, the sex of the teacher, and the maturity of the rater . . . can not be said with certainty to have any effect upon the ratings."

Remmers compared the effect of maturity in the ratings of freshmen, sophomores, juniors, and seniors. As a result of these comparisons, Remmers concluded that "the differences are relatively unimportant as compared to the resemblances among the four classes."[4]

---

[1] H. H. Remmers, "The Relationships Between Students' Marks and Student Attitude Towards Instructors," School and Society, XXVIII (December 15, 1928), 759-60.

[2] Earl Hudelson, "The Validity of Student Ratings of Instructors," School and Society, LXXIII (April 28, 1951), 265-66.

[3] J. D. Heilman and W. D. Armentrout, "The Rating of College Teachers on Ten Traits by Their Students," Journal of Educational Psychology, XXVII (March, 1936), 197-216.

[4] H. H. Remmers, L. Hadley, and J. K. Long, "Learning Effort and Attitude as Affected by Class Size in Beginning College Mathematics," Purdue Studies in Higher Education XIX, XXXII, No. 9, 44-45.

In addition, Amotora[1] found that even elementary students give rather stable ratings; also these students evidence good discrimination and agreement. It might be well to note that Symonds[2] found pupil ratings correlate positively with principal ratings for the same teachers.

Boardman[3] found a reliability of .81 for pupil rankings against .88 for supervisor ranking of these same teachers. Davenport, obtaining similar reliabilities, concluded that "it can be said with a fair degree of confidence that pupils are competent to rate teachers and that their ratings are reliable and valid, and that the ratings of pupils have no deleterious effects on either pupil or teacher morale."[4] The author also noted that pupils are the only competent judges of how much they like teachers. They are capable of rating the frequency of teaching practices, and that while such ratings are subjective, that pupils form opinions quickly and do not tend to change them.

Douglas[5] pointed out that student opinion is reliable. He reported in one experiment a coefficient of .89

---

[1] Mary Amotora, "Teacher Rating by Younger Pupils," Journal of Teacher Education, V (June, 1954), 149-52.

[2] Percival M. Symonds, "Characteristics of the Effective Teacher Based on Pupil Evaluation," Journal of Experimental Education, XXIII (June, 1955), 239-310.

[3] C. W. Boardman, "An Analysis of Pupil Rating of High School Teachers," Educational Administration and Supervision, XVI (September, 1930), 440-46.

[4] Kenneth Davenport, "An Investigation of Pupil Ratings of Certain Teaching Practices," Studies in Higher Education, XLIX (Purdue University, January, 1944), 12.

[5] Harl R. Douglas, "Rating the Effectiveness of College Instructors," School and Society, XXVIII (August 18, 1928), 192-97.

between two sets of student ratings with a one month spread between ratings. Detchen also found "consistent agreement among the ratings of thirty-eight instructors by their classes."[1]

Summary

It can be seen from this review of the literature that students can make a contribution by providing informational feedback to their teachers. This information can be very valuable to the teacher since the primary test of a teacher's effectiveness is the impact he has upon students. The best teacher is the one who brings about desirable changes in a large percentage of his students. The students in this study spend half of each day with their vocational teachers and consequently they are in a position to know much about the teachers and their behavior. There is a great body of evidence to support the fact that students do know whether they are working or loafing, whether they are confused or working with a clear purpose, and whether they are inspired or bored.

Experiments conducted also indicate that: (1) Students are keen judges of teaching; (2) As pupils become older they change relatively little in their reactions to teachers; (3) Pupil ratings on most items are highly reliable; (4) The ratings of 25 pupils whether in junior or senior high school will produce reliabilities of .90 or above on a majority of items which is as reliable as the best standardized mental and educational tests available; (5) Pupils showed much more discrimination in their ratings than did administrators; (6) Pupil ratings can be both valid and reliable measures of pupil opinion if scientically gathered; (7) There is a very low correlation between students' marks, class size, sex, severity of grading, sex of teacher, maturity of the rater, and the ratings gained by students; (8) Student ratings have no deleterious effects upon either pupil or teacher morale; (9) Years of teaching experience and age must be taken into account as a relevant independent variable whenever teacher characteristics are considered; (10) The image of a teacher held by students usually has much in common with the image held by administrators and parents; (11) The image one group of students has of a teacher is usually very similar to that held by other groups; (12) A large percentage of high school teachers can use information gained from student reaction feedback as a means of improving their image with their students; and (13) Students do not necessarily have to know what constitutes good teaching in order to furnish valuable evidence in well designed studies.

---

[1] Lily Detchen, "Shall the Student Rate the Professor?" Journal of Higher Education, XI (March, 1940), 151.

While this study borrows much from the studies reviewed in this chapter, it carries the area of research forward in at least two new directions. It is unique in that none of the studies reported thus far have sought to compare the differential effectiveness of informational feedback as a function of source. In the present research the effectiveness of student feedback was compared to that of supervisory feedback using the same opinion-questionnaire. The effect of combined student and supervisory reactions were also investigated. The 12-week interval was different from that used in previous studies. While the earlier researchers used relatively short feedback intervals varying from a few days to less than two months, later studies used one year intervals; this study used three months. This is also the first student reaction study reported to use vocational instructors. Since most vocational shop teachers came directly from industry with little previous college training, the study holds promise of making a contribution to their professional improvement. Also, the differential effectiveness of feedback from various sources is being examined for teachers of different experience groups to provide unique information on experience-source interactions.

CHAPTER III

DESIGN OF THE STUDY

In this chapter the general design of the study is presented and procedures for selecting samples and collecting data are described. The instruments used for collecting data from pupils and supervisors are discussed, as are the procedures for communicating informational feedback to the instructors. The manipulation of the independent variable and the measurement of the dependent variables are also discussed. Finally, the procedure for analyzing the data is outlined.

Specific Statement of the Problem

Specifically, this study sought to test three hypotheses. The first hypothesis was that teachers exposed to informational feedback would change more than those that are not so exposed.

Operationally stated, the mean difference scores, i.e. the differences between posttest and pretest student judgments for each of the experimental groups:

A. Those receiving student feedback.

B. Those receiving supervisor feedback.

C. Those receiving both student and supervisory feedback.

Will change more than the control group, i.e. those receiving no informational feedback. (The no-feedback group will hereafter be referred to as the control group.)

The second hypothesis was that teachers exposed to different sources of informational feedback would not change to the same degree, i.e. the sources have a different effect.

The mean difference scores[1] of the three experimental groups:

A. Supervisor only

B. Students only

C. Supervisor plus students

would not be equivalent.

The final question to be answered was: Would there be a relationship between the number of years a teacher had been teaching and his receptivity to feedback? It was hypothesized that the amount a teacher changed as a result of informational feedback would be inversely related to the amount of his teaching experience.

Operationally, the third hypothesis stated that the mean difference scores for beginning teachers (1-3 years) would exceed those having 4-10 years of experience which in turn would exceed those having 11 years or more. Finally, the beginning teachers would also exceed those having 11 years or more.

The final question to be answered was: Would there be a relationship between the interval of feedback and the amount of change in the scores at posttest? It was hypothesized that the longer interval would result in more change than a shorter interval.

---

[1] The difference between the posttest scores and the pretest scores on a student opinion-questionnaire analyzed by analysis of variance using the unweighted means method, will hereafter be referred to as the mean difference scores.

Operationally, the fourth hypothesis stated that the mean difference scores for teachers receiving informational feedback at 12 and 52 weeks from pretest would exceed those of teachers receiving feedback at 40 weeks which in turn would exceed those receiving feedback at 12 weeks.

General Design

The 2 X 2 X 3 factorial design used in this study appears in Figure 1. Instructors were separated into three groups based upon the number of years of teaching experience and then randomly assigned to one of four feedback conditions: (1) supervisor only, (2) students only, (3) supervisor and students combined, and (4) a control group which received no feedback.

The dependent variable was the amount of change in students' opinions on an opinion-questionnaire before and after treatment.

The Sample

In selecting subjects for this study the population constituted those vocational-technical instructors teaching in public schools and technical institutes in New Jersey; Wilmington, Delaware; Bucks County, Pennsylvania; and Rockland County, New York. The population from which the sample was drawn was limited to vocational instructors in the fields of vocational agriculture, distributive education, and trade and technical education. A majority of the instructors in the sample were from the trade and technical fields.

The grade levels of students from which data were collected included 10, 11, and 13. Upper classmen were excluded. Only vocational skill subject instructors were used in this study, since they had a longer class period contact with the students than did other instructors. This factor tended to lessen the chance that the phenomena under study would be unduly attenuated by the pupils' interactions

FIGURE 1

ASSIGNMENT OF INSTRUCTORS TO TREATMENT AND EXPERIENCE GROUPS

	Years of Experience of Instructor					
	1-3 years A_1		4-10 years A_2		11 or more years A_3	
	No Superv. Feedback B_1	Superv. Feedback B_2	No Superv. Feedback B_1	Superv. Feedback B_2	No Superv. Feedback B_1	Superv. Feedback B_2
No Student Feedback C_1	14*	18	19	18	18	13
Student Feedback C_2	39	32	25	31	32	27

*Cell entries are number of observations per cell. Total N=286 with other instructors. The subjects of the study included 286 vocational-technical instructors.

The median class size was 15 in both control and experimental groups and the quartile values were similar. There were 226 male and 60 female instructors included in the study. Schools where the control and experimental instructors were located are shown in Appendix M.

It is apparent that the schools are well distributed over the areas of the state in concentrations somewhat similar to those of the total population. Ages of instructors are shown in Appendix L. The mean age was 43.7 years, and the median 45.

Due to the cooperation received from all the schools contacted, the sample represented essentially all of the vocational instructors in those schools participating. In all but one county system the instructor cooperation was 100 per cent. The schools involved in the study are listed in Appendix K.

An attempt was made to retain every possible instructor and his class for the entire research study. Only three instructors of 289 whose classes were pretested were not available for the posttest.

Variables such as class size and geographic location were probably equally distributed among the groups.

Analysis of variance by the unweighted means method was used to account for the possibility that teachers may have differed at the time of pretest.

Sources of Informational Feedback

The sources of feedback were: supervisor only, student only, and combined supervisor and students.

Form 6 of the Student Opinion-Questionnaire developed by Roy C. Bryan was administered to all student, supervisor, student and supervisor combined groups, and the control group at the beginning of the study. (Psychometric characteristics of this instrument will be described in more detail in the next section.)

The experimental group receiving informational feedback from the supervisor only, received a summary of the reactions of his supervisor to the 10 items on the front page of the instrument in graphic form. This information indicated the extent to which his supervisor felt he was performing in the shop relative to items on the Opinion-Questionnaire. (See Appendix D.) The summary of the comments on the back of supervisor-opinion questionnaire was also included in the supervisory informational feedback. (See Appendix C.)

The student only experimental group received the feedback of students' reactions in graphic form and a summary of student comments.

The last group received both supervisory and student reaction reports, as well as the summary of comments from both sources. These were given to the teacher labeled as to source, i.e. they were not pooled.

Years of Teaching Experience

The independent variable, years of teaching experience was determined by the use of a personal data form as shown in Appendix E. As the completed student opinion-questionnaires came in from the schools, each teacher was located as falling into the 1-3 years experience group, the 4-10 years group, or the 11 or more years group. Teachers in the three experience groups were then randomly assigned to one of the informational feedback conditions.

Measurement of the Dependent Variable

The dependent variable was the amount of change in students' ratings of their instructor on the Student Opinion-Questionnaire before and after treatment. For all treatment conditions, the dependent measure was based on the students' measure of change between posttest and pretest.[1] The scores given by each student on the ten questionnaire items were averaged across all the students in the class on an item-by-item basis to get a single pre-treatment and post-treatment score on an instructor (the instructor being the unit of analysis) for each item.

Student Opinion-Questionnaire and Data Collection

The principal considerations for determining the selection of the instrument used to collect ratings from students and supervisors in this study briefly are:

A. Items brief and defined objectively.

B. Numbers of items small enough to be answered within one class period.

C. Items describe recognizable behavior.

D. Number of steps in the scale limited to 5.

E. Scale can be agreed upon by competent judges.

F. Provide an opportunity for favorable and unfavorable comments.

G. High reliability coefficients for the different items.

---

[1] Student judgments were used for measurement purposes (even when the feedback came from supervisors) since a) there are more students serving as observers, b) students observe the teacher more often than supervisors, and c) in order to keep the measuring instrument constant across conditions. Since the pretest means of the group rated by supervisor only and the group rated by students only, did not differ significantly when tested by t-test, these two sets of scores can be considered interchangeable and permit the use of student judgments as the criterion measure for all treatment groups.

(They appear in more detail in Appendix A.) In the light of these outlined considerations, the instrument developed by Roy C. Bryan for use in his Student Reaction Center at Western Michigan University was selected, in that it met the above criteria to a high degree.

The reliability coefficients for the different questions range from .75 ± .06 to .85 ± .04 when chance-half averages for 50 classes (50 teachers) are correlated. Bryan[1] found that for whole classes numbering 28 students on the average, the coefficients range from .86 ± .04 to .92 ± .02. The reliabilities reported by Bryan were by students in Michigan in grades 7 through 12.

The opinions of students were obtained by means of the Student Opinion-Questionnaire in Appendix B. The administrators were asked to rate the instructors using the same kind of rating sheets that were used by the pupils with one exception. The wording of Item 6 was changed to read: How much are the pupils learning in this class? A copy of the memorandum sent to the administrators along with the Opinion-Questionnaires is shown in Appendix H.

In most of the schools included in this study, the vocational guidance counselor acted as the coordinator. This worked especially well since the counselor had the confidence of the pupils, the instructors, and the administrators. Special instructions for the Administration of the Opinion-Questionnaire are contained in Appendix G.

Communicating Feedback

After the Opinion-Questionnaires were completed, they were mailed to the investigator by the vocational guidance counselor from each school.

---

[1] Roy C. Bryan, Reactions to Teachers by Students, Parents, and Administrators, United States Office of Education, Cooperative Research Project No. 668 (Kalamazoo, Michigan: Western Michigan University, July 1, 1959), p. 44.

An IBM card for each student was programmed for the computer and a "run" sheet was obtained giving the mean for each of the 10 items for each instructor's class. These means were plotted on a Student Reaction Report Form shown in Appendix D. A summary of comments for an instructor is shown in Appendix C.

Those instructors assigned to the experimental groups were sent the summaries on questions 11-14, a Student Reaction Report Form similar to that in Appendix D and the memorandum contained in Appendix F.

The date that the feedback was mailed was recorded in the log book and 12 weeks later the posttest was scheduled and administered. Instructors in the control group were told that their reports were being delayed and would be mailed to them after the posttest was administered. Students were not given information as to which treatment group their reactions would be included in. They were assured that the study was designed to help their instructor increase his effectiveness and that he would be the only person to see the student reactions other than the researcher.

Analysis of Data

The data was analyzed by a 2 X 2 X 3 factorial experiment analysis using the unweighted means method of analysis. This method was used in order to take care of unequal cell frequencies.[1]

---

[1] Loss of observations in cells is essentially random, therefore, the experimental data may be analyzed appropriately by analysis of variances by the unweighted means method. In essence this method treats each cell as if it contained the same number of observations as all the other cells with regard to the computation of main effects and interaction effects. Each cell is considered to have a number of observations per cell. For this factorial experiment the harmonic mean is .4736. See B. J. Winer, Statistical Principles in Experimental Design (New York: McGraw-Hill, 1962), p. 222.

Differences in the mean difference scores were compared by the Least Significant Difference Method. The least significant difference score was computed for each item by using the following formula:

$$LSD_2 = t_2 \, S_d = t_2 \sqrt{\frac{2S^2}{N}}$$

A discussion of this method is shown in Appendix J.

All analysis was performed by the 7040 computer in the Rutgers Data Processing Center.

Steps in the analysis of the data were as follows:

1. The difference between posttest and pretest means on each item was obtained for each instructor.

2. A mean difference score on each item was obtained for each of the 286 instructors (i.e. Posttest-Pretest difference). These instructors were grouped according to type of feedback. The stratification included control groups which received no feedback and experimental groups with feedback from supervisors only, students only, and both supervisors and students. These groups were further stratified by years of teaching experience with experience Group 1 being those with 1-3 years of teaching experience; Group 2, 4-10 years of teaching experience; and Group 3 with 11 years and over. The above groups were randomly assigned to treatment groups to form a 2 X 2 X 3 factorial experiment.

3. Analysis of variance using the unweighted means was used to obtain mean difference scores between posttest and pretest for each of the treatment groups on each of the 10 items in the student opinion questionnaire.

4. Significant differences between the mean difference scores for each treatment group were tested for significance at the .05 level by the Least Significant Difference method.

CHAPTER IV

PRESENTATION OF DATA

The purpose of this chapter is to present the results of the investigation. The same subjects (teachers, supervisors, and pupils) were not involved in all of the items, therefore, in presenting the results the 10 items in the rating scale are treated independently. The results are organized under three major headings: (1) the control group compared to the three experimental groups, (2) differences related to source of informational feedback, and (3) differences related to the number of years of teaching experience.

The first concern of the study was to answer the question, "Would teachers, as rated by their pupils and/or supervisors, change more if they were given information about how their pupils and/or supervisors described them than if not given this information?" Teachers in experimental groups were given this information, while those in the control group were not. A second question to be answered by the study was, "Were there significant differences caused by the sources of informational feedback?" The final question to be answered by the study was, "Would teachers who have few years of teaching experience change more than those with a greater number of years of experience?"

Differences Between the Control and Experimental Groups

The first hypothesis tested was: There are significant differences between pretest and posttest scores (i.e., mean difference scores) obtained in the student opinion-questionnaire between the control group and the three experimental groups. Data bearing on this hypothesis are shown in Tables 1 and 2.

TABLE 1

ANALYSIS OF VARIANCE OF ITEMS 1-5
(A) TEACHER EXPERIENCE, (B) SUPERVISOR REACTIONS AND (C) STUDENT REACTIONS

Source	d.f.	1 Knowledge		2 Explanation		3 Fairness		4 Discipline		5 Sympathy	
		MS	F	MS	F	MS	F	MS	F	MS	F
FACTOR A Teacher Experience	2	.007	1	.005	1	.004	1	.015	1	.001	1
FACTOR B Supervisor Reaction	1	.011	1.29	.011	.786	.037	2.191	.006	1	.017	1.041
FACTOR C Student Reaction	1	.013	1.58	.110	8.22[b]	.058	3.471[c]	.102	6.55[b]	.014	.741
AB	2	.004	1	.012	1	.015	1	.011	1	.0011	1
AC	2	.003	1	.053	3.95[a]	.003	1	.004	1	.007	1
BC	1	.005	1	.0001	1	.018	1.08	.0001	1	.030	1.82
ABC	2	.003	1	.003	1	.015	1	.006	1	.005	1
TOTAL ERROR	274	.008		.0134		.0168		.0159		.0167	

[a] .05 Critical Value $F_{.95}$ (2,274) = 3.00
[b] .05 Critical Value $F_{.95}$ (1,274) = 3.84
[c] .10 Critical Value $F_{.90}$ (1,274) = 2.71

TABLE 2

ANALYSIS OF VARIANCE ON ITEMS 6-10
(A) TEACHER EXPERIENCE, (B) SUPERVISOR REACTIONS AND (C) STUDENT REACTIONS

Source	d.f.	Amount of Learning		Interesting		Business-like		Think		All-round Ability	
		MS	F	MS	F	MS	F	MS	F	MS	F
FACTOR A Teacher Experience	2	.058	1	.018	1.12	.002	1	.016	1	.017	1.12
FACTOR B Supervisor Reaction	1	.026	1	.010	1	.035	2.34	.039	1.97	.025	1.61
FACTOR C Student Reaction	1	.741	11.0^b	.027	1.09	.028	1.88	.064	3.46^c	.055	3.51^c
AB	2	.055	1	.004	1	.025	1.58	.002	1	.007	1
AC	2	.067	1	.008	1	.020	1.34	.003	1	.032	2.03
BC	1	.010	1	.007	1	.003	1	.019	1	.010	1
ABC	2	.034	1	.002	1	.072	4.79^a	.012	1	.030	1.92
TOTAL ERROR		.068		.0159		.015		.018		.016	

^{a}.05 Critical Value $F_{.95}$ (2,274) = 3.00
^{b}.05 Critical Value $F_{.95}$ (1,274) = 3.84
^{c}.10 Critical Value $F_{.90}$ (2,274) = 2.71

The main effect of supervisory feedback (B) was not significant on any item indicating a lack of difference between it and the control group. The main effect of student feedback (C) was significant on items 2, 4, and 6 and approached significance on items 3, 9, and 10 indicating a difference from the control group.

From the differences in means and least significant difference scores shown in Table 3 effects could be further pinpointed. It was found that the Supervisor Only group did not differ significantly from the control group on any item. The group receiving informational feedback from Students Only differed significantly at the .05 level from the control group on Item 2 (Explanations), Item 4 (Discipline), Item 6 (Amount of Learning), and Item 9 (Think). The group receiving informational feedback from Students and Supervisors combined[1] showed increases when compared to the control group which were significant at the .05 level on Item 2 (Explanations), Item 4 (Discipline), and Item 6 (Amount of Learning).

Overall, therefore, it was found that the mean difference scores for two of the experimental groups differed significantly at the .05 level from the mean difference scores of the control group on at least 3 items. Only the group receiving informational feedback from the supervisors only did not show a significant change on any item when compared to the control group.

These findings confirm Hypothesis 1 as regards Student Only and Students plus Supervisor feedback sources when compared to the control group. Hypothesis 1 is rejected as regards differences between the Supervisor Only group and the Control group.

---

[1] Because supervisory and student feedback were treated as separate factors in the analysis of variance, any effect of these two sources of effect, if not beyond that of either alone, could not be detected in the analysis of variance. However, a comparison of treatment means as shown in Table 3 makes it possible to separate out the joint effect of the two feedback sources.

TABLE 3

MEAN DIFFERENCE SCORES CALCULATED BY ANALYSIS OF VARIANCE FOR 10 ITEMS USED TO EVALUATE VOCATIONAL TEACHERS BY THREE GROUPS OF RATERS FOR A CONTROL AND THREE EXPERIMENTAL GROUPS

Item	Least Significant Difference	Control	Experimental Groups		
			Supervisor Only	Students Only	Students and Supervisor Combined
1. Knowledge	.092	-0.10748	-0.16489	-0.09399	-0.11158
2. Explanations	.092	-0.15489	-0.21778	+0.03322[a]	-0.02245[a]
3. Fairness	.216	-0.16393	-0.35262	-0.10247	-0.13520
4. Discipline	.020	-0.21219	-0.26628	-0.03708[a]	-0.07221[a]
5. Sympathy	.040	-0.12128	-0.29795	-0.15283	-0.012801
6. Amount of Learning	.298	-0.06052	-0.21067	+0.27912[a]	+0.34385[a]
7. Interesting	.085	-0.14113	-0.24801	-0.09510	-0.10405
8. Businesslike	1.38	-0.14024	-0.21494	-0.01000	-0.15147
9. Think	.058	-0.06578	-0.25963	+0.00128[a]	-0.03467
10. All-round Ability	.284	-0.06694	-0.21651	+0.01040	-0.02318

[a] Differs significantly at the .05 level from the control group.

Differences Due to Source of Informational Feedback

The second hypothesis tested was: There are significant differences in the mean difference scores relative to the source of informational feedback:

A. Between students and supervisors.

B. Between students and the combined ratings of students and supervisors.

C. Between supervisor only and the combined ratings of students and supervisors.

Data bearing on this hypothesis are shown in Table 4. From the mean difference scores calculated by analysis of variance and recorded in Table 4 it was found that the posttest-pretest differences differed significantly at the .05 level for five of the 10 items when the group receiving student feedback was compared to the group receiving supervisor only feedback. The five items which showed significant differences were: Item 2 (Explanations), Item 3 (Fairness), Item 4 (Discipline), Item 6 (Amount of Learning), and Item 7 (Interesting).

No significant differences were observed, when the scores of the students only group were compared to those of the group receiving informational feedback from both students and supervisors. The scores were comparable on all 10 items.

When the mean difference scores of supervisor only were compared to students and supervisors combined, significant differences at the .05 level were revealed on Item 2 (Explanations), Item 3 (Fairness), Item 4 (Discipline), Item 6 (Amount of Learning), and Item 7 (Interesting).

These findings confirm Hypothesis 2 as regards the relative effects of the different feedback sources. The data revealed significant differences between Supervisor Only and Student Only feedback and between Supervisor Only and Combined Student and Supervisor feedback. However, differences between the Student Only group and the group

TABLE 4

MEANS CALCULATED BY ANALYSIS OF VARIANCE FOR 10 ITEMS USED TO EVALUATE VOCATIONAL TEACHERS BY THREE EXPERIMENTAL GROUPS

Item	Least Significant Difference	Mean Differences of Experimental Groups		
		Supervisor	Student Only	Student and Supervisor
1. Knowledge	.092	-0.16489	-0.09399	-0.11158
2. Explanations	.092	-0.21778	+ 03322[a]	-0.02245[a]
3. Fairness	.216	-0.35262	-0.10247[a]	-0.13520[a]
4. Discipline	.020	-0.26628	-0.03708[a]	-0.07221[a]
5. Sympathy	.040	-0.29795	-0.15283	-0.12801
6. Amount of Learning	.298	-0.21067	+0.37912[a]	+0.34385[a]
7. Interesting	.085	-0.24801	-0.09510[a]	-0.10405[a]
8. Businesslike	1.38	-0.21494	-0.01000	-0.15147
9. Think	.058	-0.25963	+0.00128	-0.03467
10. All-round Ability	.284	-0.21651	+0.01040	-0.02318

[a] Differs significantly at .05 level from Supervisor Only group.

receiving feedback from Both Students and Supervisors were not obtained.

Overall, then, the four feedback conditions can be grouped in terms of change produced in teachers as follows.

amount of change

→

no feedback (Control) Supervisors Only	Students Only Supervisors and Students Combined

The implications of this finding will be discussed in the Discussion section.

Differences Due to Years of Teaching Experience

Hypothesis 3 predicted a differential effect of informational feedback as an inverse function of a teacher's years of teaching experience. No main effect of years of experience (A) was obtained in the analysis of variance. Mean comparison data provided in Table 5 revealed the following differences.

The mean difference scores for beginning teachers (1-3 years) compared to experienced teachers with 4-10 years differed significantly at the .05 level on Item 5 (Sympathy). No significant differences were observed on any other item.

However, when the teachers with 4-10 years of experience were compared to teachers with 11 or more years of experience, significant differences at the .05 level were observed on five items. These items were: Item 2 (Explanations), Item 4 (Discipline), Item 5 (Sympathy), Item 9 (Think), and Item 10 (All-round Ability).

TABLE 5

MEANS CALCULATED BY ANALYSIS OF VARIANCE FOR 10 ITEMS USED BY VOCATIONAL TEACHERS IN THREE EXPERIENCE GROUPS

Item	Least Significant Difference	Experience Group Means		
		1 - 3	4 - 10	11 & over
1. Knowledge	.092	-0.04830	-0.03501	-0.12487
2. Explanations	.092	+0.06599[a]	+0.11045[a]	-0.16028
3. Fairness	.216	-0.12983	-0.09237	-0.13431
4. Discipline	.020	-0.04729[a]	+0.02814[a]	-0.114479
5. Sympathy	.040	-0.08128[b]	-0.14156[a]	-019841
6. Amount of Learning	.298	+29341	+0.16174	+0.62931[c]
7. Interesting	.085	-0.12559	-0.06741	-0.10572
8. Businesslike	1.38	-0.00829	-0.09198	-014193
9. Think	.958	+0.00275[a]	+06233[a]	-0.115616
10. All-round Ability	.284	+0.12002[a]	+0.12143[a]	-0.17818

[a] Signiffcantly greater than group 3 at the .05 level.

[b] Significantly greater than group 2 or 3 at the .05 level.

[c] Significantly greater than group 1 or 2 at the .05 level.

TABLE 6

MEAN DIFFERENCE SCORES CALCULATED BY ANALYSIS OF VARIANCE FOR 10 ITEMS USED TO EVALUATE VOCATIONAL TEACHERS BY STUDENT RATERS USING TWO TIME INTERVALS (12 WEEKS AND 40 WEEKS)

Item	Least Significant Difference	Time Intervals Between Pretest and Posttest		
		12 weeks N=28 Fall & Spring	12 and 40 weeks N=28 Spring & Fall Two Posttests	40 weeks N=39 Spring & Fall One Posttest
1. Knowledge	.338	-0.143	+0.019	-0.048
2. Explanations	.300	-0.209	+0.089	+0.033
3. Fairness	.374	-0.350	+0.125	-0.093
4. Discipline	.518	-0.284	+0.060	-0.010
5. Sympathy	.418	-0.346	+0.066	-0.167
6. Amount of Learning	.442	-0.213	-0.061	-0.004
7. Interesting	.328	-0.166	+0.010	-0.043
8. Businesslike	.388	-0.274	-0.051	+0.002
9. Think	.328	-0.307	+0.021	+0.009
10. All-round Ability	.328	-0.157	+0.25	+0.042

When the beginning teachers (1-3 years) were compared to those having 11 or more years of experience differences in mean difference scores between pretest and posttest for the beginning teachers were significantly greater than the experienced group (11 years or over) on Item 2 (Explanations), Item 4 (Discipline), Item 5 (Sympathy), Item 9 (Think), and Item 10 (All-round Ability).

Only on Item 6 (Amount of Learning) did the group having 11 or more years of experience show a significant gain over the beginning teachers and those with 4-10 years of experience.

Hypothesis 3 is confirmed as regards the comparison between beginning teachers and teachers having 11 or more years of experience, the former having shown greater change to feedback. It is also confirmed with regard to those in the 4-10 year experience group who made significant gains over those with 11 years of experience or more. The hypothesis is rejected, however, as regards the comparison between beginning teachers and those with 4-10 years of experience, these two groups being comparable on all items.

Hypothesis 4 is rejected as regards the comparison of a longer time interval. There were no significant differences between the group receiving feedback at 52 weeks with an additional feedback at 12 weeks and the group which received feedback after 40 weeks. Although the smallest difference was shown by the group with the 12 week interval this difference was not significant at the .05 level.

Although the group with the 40 week interval having a pretest in the spring and the posttest in the fall showed gains on all but two items these gains did not differ significantly from the other two groups.

CHAPTER V

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

The major purpose of the study was to compare the relative effectiveness of informational feedback to instructors from three sources. The method of taking account of initial differences in control and experimental groups was analysis of variance using the unweighted means. The pretest ratings served as the control variables, the posttest ratings as the dependent variables. Three levels of Factor A (Teaching Experience), two levels of Factor B (Supervisor Reactions) and two levels of Factor C (Student Reactions) served as the independent variables. The three sources of informational feedback were students only, supervisor only and both supervisor and students. The grade levels used were 10, 11 and 13. The instrument used to gather the feedback was Form 6 of the Student Opinion-Questionnaire.

A second purpose of the study was to compare the relative effectiveness of informational feedback as it related to the number of years of experience of teachers. The three experience groups compared were: beginning teachers (1-3 years), experienced teachers 4-10 years and experienced teachers with 11 years and over.

There were no significant differences due to Main Effect A (Teacher Experience) on any of the 10 items. This may have been due in part to the effect of age. Since vocational instructors must present several years of occupational experience before they begin to teach, the average age of beginning teachers is in the late thirties.

In testing Hypothesis 1 relating to significant differences between those groups receiving informational feedback and those not receiving feedback the following was revealed. Teachers receiving feedback from Supervisor Only did not differ significantly from teachers receiving no feedback on any item at the .05 level. This appears due to the fact that informational feedback was received from only one source, i.e. a single supervisor. Hence it was not nearly as complete as feedback received from 10-30 students who have intimate knowledge of what goes on in a shop or laboratory over a three hour period, daily. The supervisor has only occasional contact with the teaching situation and therefore is not in a position to furnish as much information.

The teachers receiving informational feedback from Students Only differed significantly at the .05 level from teachers receiving no feedback on Item 2 (Explanations), Item 4 (Discipline), Item 6 (Amount of Learning), and on Item 9 (Think). The teachers receiving informational feedback from the Students and Supervisors Combined differed significantly from teachers receiving no feedback at the .05 level on Item 2 (Explanations), Item 4 (Discipline), and on Item 6 (Amount of Learning).

It is interesting to note that the Students Only group and the Students and Supervisors group differed significantly from the control group on three or more items, whereas the mean difference score for the Supervisor Only group did not differ significantly from the Control group on any item. This again leads us to question the effectiveness of supervisory feedback.

These findings confirm the first hypothesis only as regards feedback from Students and Students and Supervisors combined.

In testing Hypothesis 2, as regards differences due to source of informational feedback, the data revealed the following differences: The group receiving Student Only feedback differed significantly at the .05 level from the group receiving feedback from Supervisor Only on 5 of the 10 items. The items which showed significant gains on the mean difference scores were: Item 2 (Explanations), Item 3 (Fairness), Item 4 (Discipline), Item 6 (Amount of Learning), and Item 7 (Interesting). When the mean difference scores between Posttest and Pretest were compared for the Supervisor Only group and the Supervisor plus Students group the latter group made significant gains on the same five items listed above, i.e. Items 2, 3, 4, 6 and 7.

However, when the Students Only group was compared to the group receiving informational feedback from both Students and Supervisors, no significant differences were observed on any item. The scores were comparable on all 10 items.

When the mean difference scores of the Supervisor Only group were compared to the group receiving feedback from both Students and Supervisors, findings comparable to the Supervisor Only--Student Only comparisons were obtained. Since the Student Only group and the Supervisor plus Student feedback group each differ significantly from the Supervisor Only group but not from each other we can infer that supervisor feedback makes no difference above and beyond student feedback. Hypothesis 2 is confirmed in that feedback from Students has an effect on teachers while feedback from Supervisors has no effect either by itself or as an increment above and beyond the effect of student feedback.

One is struck by the large number of negative mean difference scores in Tables 3, 4, and 5. However, when the posttest means of the control groups were compared by t-test to the scores of the posttest only group, i.e. the group which had no pretest--only a posttest, the means did not differ significantly. Since there were significant differences at the time of posttest between the control group and the other treatment groups, and no significant difference between control group and posttest only group, one is led to infer that the time of year during which data is collected may have some effect. It appears that students in general tend to be more critical in their reactions to teacher behavior in the Spring near the end of the year, than they are in the Fall when the school year is starting. In the other studies using this instrument, data was collected only in the Spring of the year. One can further conclude that negative changes were not the result of a sensitizing effect from the pretest since posttest scores of the control group and posttest only group were equivalent.

In testing Hypothesis 3 as regards the differential effect of informational feedback over number of years of teaching experience the data revealed the following differences.

The differences in the mean difference scores for beginning teachers (1-3 years) showed significant gains at the .05 level when compared to the 4-10 year experience group on one item, and on five items when compared to the 11 or more year experience group.

When the mean difference scores for teachers with 4-10 years of experience were compared to those having 11 or more years, significant gains were observed on the same five items as differentiated between most and least experienced groups.

Interactions between years of experience and source of feedback were generally not obtained. Hypothesis 3 was generally confirmed in that sensitivity to feedback and years of teaching experience were inversely related. This inverse relationship holds most closely if 1-3 and 4-10 year groups are combined and contrasted with 11 or more year group. Perhaps a critical period occurs around seven years during which time and insensitivity to feedback emerges.

It appears that the experience groups for beginning teachers were too narrow to allow for adequate discrimination relative to sensitivity to feedback. The rationale for using the one to three years experience grouping was that the beginning of the fourth year is the point where most teachers are given tenure. The vocational teacher differs from the academic teacher with respect to the age of entry into teaching. Most academic teachers enter teaching directly upon graduation from college at the age of 21 or 22. Since vocational teachers are required to have occupational experience before becoming eligible to teach, only a small number are employed before the age of 30. Of the 286 vocational teachers included in this study, only 17 were under the age of 30. In the age group 35 to 50, 179 instructors were included. In Appendix L it can be seen that a majority of the vocational teachers fall in the 35 to 50 year age group. One concludes that a wider spread for the beginning teacher experience group is needed in order to differentiate differences in sensitivity to informational feedback as it relates to years of teaching experience.

Since vocational teachers tend to enter teaching during their middle thirties to early forties, there is a tendency for the most experienced teachers to be well into their fifties. One wonders whether age may not be a significant factor which has a greater effect than the number of years of teaching experience. In the review of the literature section, many studies were reviewed which tended to conclude that age must be considered as a relevant variable whenever teacher characteristics are considered. It is significant, however, that the experienced teachers with 11 or more years did make significant gains on one very important item, namely Item 6, Amount of Learning.

Differences due to the interval of feedback were generally not obtained. Hypothesis 4 was rejected in that there were no significant differences due to the feedback interval. It appears that the results may have been confounded somewhat by the time of year during which the data was gathered. The 40 week interval had the pretest in the spring and the posttest in the fall whereas the 12 week interval had the pretest in the fall and the posttest in the spring. At the time the study was designed there was no way to anticipate that there would be an effect due to when the data was collected. In order to make a valid comparison the time intervals should have covered the same semesters.

Implications

The results of this study have the following implications:

1. Informational feedback from students is effective in changing teacher behavior.

2. Student feedback is more effective in changing teacher behavior than is supervisory feedback.

3. The utilization of student feedback as a means of influencing teacher behavior should be used to a greater extent. (Except for the work of Roy C. Bryan at Western Michigan and Dwight Allen, et al., at Stanford University, little is being done at the present time to exploit this very important source of aid for the teacher who is seeking to improve his professional growth.)

Due to innumerable causes, such as lack of time and adequate staffing, little progress has been made in American education regarding the fundamental task of helping beginning classroom teachers overcome their initial deficiencies. Informational feedback from students, under appropriate conditions, promises to be an economical and effective means of helping new teachers improve their effectiveness. This task could best be implemented by establishing a Student Reaction Center similar to the one directed by Roy C. Bryan at Western Michigan University. By further refining the Student Opinion Questionnaire, so as to permit the use of an optical scanning system of handling the data, this could be accomplished at a minimum cost. The cooperation received from the teachers in this study reveals a willingness upon the part of most instructors to participate in a program which gives promise of helping them to do a more effective job of teaching.

The use of student feedback could be most helpful as a part of every supervised teaching program for in-service teachers. The lack of adequate time and supervisory personnel makes it mandatory that new avenues be explored. With further refinement, this technique offers the possibility of giving the supervisor some measure of the professional growth of the teacher during his intern period. By making a more adequate determination of strengths and weaknesses of the beginning teacher, much more effective professional growth should result.

Conclusions

The findings of this study showed the following:

1. There were significant differences between those groups receiving feedback from Students, either alone or in

combination with feedback from Supervisors, and those who received no feedback.

2. The different sources of informational feedback were not equally effective. Student feedback effected teachers while Supervisor feedback did not. The effect of the two combined did not exceed that of Student feedback alone.

Based upon these findings, student feedback during the first 10 years of teaching can be used as an effective method of improving teacher effectiveness as seen by students.

Recommendations for Further Study

Recommendations made as a result of this study are:

1. That informational feedback about student reactions be used to identify areas where new teachers need assistance in their orientation to the teaching profession.

2. That a study be carried out with extended time intervals to determine what interval of feedback is most effective.

3. That further experiments be carried out to determine whether students tend to rate their instructors more favorably early in the school year than towards the end of the spring semester.

4. That further studies be carried out to determine whether or not age may be a significant factor in the instructors' reaction to informational feedback about his students reactions.

5. That further studies be carried out using a wider range of experience categories in order to determine sharper differences due to years of teaching experience.

2

BIBLIOGRAPHY

Amotora, Mary. "Teacher Rating by Younger Pupils," Journal of Teacher Education, V (June, 1954), 149-52.

Armentrout, W. D. "Improving College Teaching by Consulting Consumers," School Executive Magazine, LI (July, 1932), 476-77.

Barr, A. S. Characteristic Differences of Good and Poor Teachers. Bloomington, Illinois: Public School Publishing Company, 1929.

Beecher, Dwight. The Evaluation of Teaching. Syracuse, New York: University Press, 1949.

Boardman, C. W. "An Analysis of Pupil Rating of High School Teachers," Educational Administration and Supervision, XVI (September, 1930), 440-46.

Bowman, Earl C. "Pupil Rating of Student-Teachers," Educational Administration and Supervision, XX (February, 1934), 141-47.

Bryan, Roy C. "Opinions of Teachers Held by Former Pupils," School Review, LI (May, 1944), 555-61.

________. "Pupil Rating of Secondary School Teachers," Contributions to Education, No. 708. New York: Bureau of Publications, Teachers College, Columbia University (1937), 83-93.

________. "Reactions to Teachers by Students, Parents, and Administrators," United States Office of Education, Cooperative Research Project No. 668. Kalamazoo, Michigan: Western Michigan University (1963), 26-32.

Buellesfield, Henry. "Causes of Failure Among Teachers," Educational Administration and Supervision, I (September, 1915), 439-52.

Davenport, Kenneth. "An Investigation of Pupil Ratings of Certain Teaching Practices," Studies in Higher Education, XLIX (January, 1944), Purdue University, 12.

Davis, C. O. "The High School as Judged by Students," North Central Association of Colleges and Secondary Schools (1924).

Detchen, Lily. "Shall the Student Rate the Professor?" Journal of Higher Education, XI (March, 1940), 151.

Douglas, Harl R. "Rating the Effectiveness of College Instructors," School and Society, XXVIII (August 18, 1928), 192-97.

Druckers, A. J., and Remmers, H. H. "Do Alumni and Teachers Differ in Their Attitudes Toward Instructors?" Journal of Educational Psychology, IL (March, 1951), 129-43.

Heilman, J. D., and Armentrout, W. D. "The Rating of College Teachers on Ten Traits by Their Students," Journal of Educational Psychology, XXVII (March, 1936), 197-216.

Hudelson, Earl. "The Validity of Student Ratings of Instructors," School and Society, LXXIII (April 28, 1951), 265-66.

Knight, F. B. "Qualities Related to Success in Teaching," Teachers College Contributions to Education, No. 120. New York: Columbia University, pp. 1-3.

Leeds, Carroll H. "A Scale for Measuring Teacher-Pupil Attitudes and Teacher-Pupil Rapport," Psychological Monographs, LXIV, No. 6 (1950), 1-24.

Littler, Sherman. "Why Teachers Fail," Home and School Education (1914), 235.

MacDonald, Marion E. "Student Opinion as Regards Desirable and Undesirable Qualifications and Practices in Teacher Training Institutions," Educational Administration and Supervision, XVII (February, 1931), 139-46.

McCall, W. A. Measurement of Teacher Merit, Bulletin No. 284 (Raleigh, North Carolina: State Department of Instruction, 1959), 29 pp.

Merriam, J. L. "Normal School Education and Teaching Efficiency," Teachers College Contributions to Education, No. 1. New York: Columbia University (1905), pp. 51-103.

Morsh, Joseph E., and Wilder, Eleanor W. "Identifying the Effective Instructor: A Review of the Quantitative Studies, 1900-1950," Air Force Personnel and Training Research Center, Research Bulletin, TR-54-44. San Antonio, Texas: Lackland A.F.B., p. 39.

Peterson, Warren A. "Age, Teachers Role, and the Institutional Setting," Contemporary Research on Teacher Effectiveness. Edited by Bruce J. Biddle and William J. Ellena. New York: Holt, Rinehart and Winston, Inc., 1964.

Remmers, Hermann H. "The College Professor as the Student Sees Him," Studies in Higher Education, XI, Purdue University (March, 1929), 21.

________. "Rating Methods in Research on Teaching," Handbook of Research on Teaching. Edited by N. L. Gage. Chicago: Rand McNally and Co., 1963.

Remmers, H. H. "The Relationships Between Students' Marks and Student Attitude Towards Instructors," School and Society, XXVIII (December 15, 1928), 759-60.

Remmers, H. H., Hadley, L., and Long, J. K. "Learning Effort and Attitude as Affected by Class Size in Beginning College Mathematics," Purdue Studies in Higher Education XIX, XXXII, No. 9, 44-45.

Riley, John W., Ryan, Bryce F., and Lifshitz, Marcia. The Student Looks at His Teacher. New Brunswick, New Jersey: Rutgers University Press, 1950.

Ryans, David G. Characteristics of Teachers. Washington, D. C.: American Council on Teacher Education, 1960.

________. "Research on Teacher Behavior in Context of the Teacher Characteristics Study," Contemporary Research on Teacher Effectiveness. Edited by Bruce J. Biddle and William J. Ellena. New York: Holt, Rinehart and Winston, Inc., 1964.

Savage, Marjorie. "Changes in Student Teachers Through the Use of Pupil-Ratings." Unpublished Ed.D. dissertation, University of Illinois, 1957.

Shaie, Warner K. "Rigidity--Flexibility and Intelligence: A Cross-Sectional Study of Adult Life Span From 20 to 70 Years," Psychological Monographs, LXXII, No. 9 (1958), 261.

Starrak, J. A. "Student Rating of Instruction," Journal of Higher Education, V (February, 1934), 88-90.

Symonds, Percival M. "Characteristics of the Effective Teacher Based on Pupil Evaluations," Journal of Experimental Education, XXIII (June, 1955), 289-310.

Swineford, Edwin J. "A Study of Factors that Affect Teaching Behavior," California Journal of Educational Research, XIV (1963), 214-24.

Turner, R. L. "Teaching as Problem Solving Behavior," Contemporary Research on Teacher Effectiveness. Edited by Bruce J. Biddle and William J. Ellena. New York: Holt, Rinehart and Winston, Inc., 1964.

Ward, William D., Remmers, Hermann H., and Schmalzried, Newell T. "The Training of Teacher-Personality by Means of Student Ratings," School and Society (1941), 189-93.

APPENDIX A

CRITERIA FOR SELECTION
OF OPINION-QUESTIONNAIRE

1. The items should be brief, defined objectly and simply so that the same interpretation will be perceived by all raters.

2. The items should call for observations that high school and technical institute pupils are prepared to give in accordance with their general backgrounds.

3. The number of items included in the instrument must be small enough to enable pupils to complete the scale within one class period without feeling hurried.

4. Each item should describe a recognizable teacher behavior upon which pupil agreement is possible and highly probable.

5. The number of steps (or degrees) in the scale should be limited to five so they can be clearly defined.

6. All items in the scale should be generally agreed upon by competent judges as important in determining teaching success.

7. The instrument used to obtain pupil ratings of teachers should include a section which directs the raters to give in their own words favorable and unfavorable comments relating to points not specifically covered by the series of items in the instrument.

8. The instrument should produce: (A) higher reliability coefficients when chance half average ratings of different teachers on the same item are correlated;

(B) relatively high correlation with ratings on the item of general teaching ability; and (C) relatively low correlation when ratings on the different pairs of items are correlated except for the item on general ability.[1]

---

[1] Roy C. Bryan, Pupil Rating of Secondary School Teachers, Teachers College, Columbia University Contributions to Education, No. 708 (New York: Bureau of Publications, Teachers College, Columbia University, 1937), p. 92.

APPENDIX B

STUDENT-OPINION QUESTIONNAIRE

(Form Six)

Please answer the following questions honestly and frankly. Do not give your name. To encourage you to be frank, your regular teacher should be absent from the classroom while these questions are being answered. Neither your teacher nor anyone else at your school will ever see your answers.

The person who is temporarily in charge of your class will, during this period, collect all reports and seal them in an envelope addressed to Rutgers University. Your teacher will receive from the university a summary of the answers by the students in your class. The University will mail this summary to no one except your teacher unless requested to do so by your teacher.

After completing this report, sit quietly or study until all students have completed their reports. There should be no talking.

Underline your answer to each question on this page. Write your answers to questions 11 to 14 on the other side of this page.

WHAT IS YOUR OPINION CONCERNING:

1. THE KNOWLEDGE THIS TEACHER HAS OF THE SUBJECT TAUGHT?
(Has he a thorough knowledge and understanding of his teaching field?)
Below Average   Average   Good   Very Good   The Very Best

2. THE ABILITY OF THIS TEACHER TO EXPLAIN CLEARLY?
(Are assignments and explanations clear and definite?)
Below Average   Average   Good   Very Good   The Very Best

3. THIS TEACHER'S FAIRNESS IN DEALING WITH STUDENTS?
(Is he fair and impartial in treatment of all students?)
Below Average   Average   Good   Very Good   The Very Best

4. THE ABILITY OF THIS TEACHER TO MAINTAIN GOOD DISCIPLINE?
(Does he keep good control of the class without being harsh? Is he firm but fair?)
Below Average   Average   Good   Very Good   The Very Best

5 THE SYMPATHETIC UNDERSTANDING SHOWN BY THIS TEACHER?
(Is he patient, friendly, considerate, and helpful?)
Below Average   Average   Good   Very Good   The Very Best

6. HOW MUCH YOU ARE LEARNING IN THIS CLASS?
(Are you learning well and much? Are you really working?)
Below Average   Average   Good   Very Good   The Very Best

7. THE ABILITY THIS TEACHER HAS TO MAKE CLASSES INTERESTING?
(Does he show enthusiasm and a sense of humor? Does he vary teaching procedures?)
Below Average   Average   Good   Very Good   The Very Best

8. THE ABILITY OF THIS TEACHER TO GET THINGS DONE IN AN EFFICIENT AND BUSINESS-LIKE MANNER?
(Are plans well made? Is little time wasted?)
Below Average   Average   Good   Very Good   The Very Best

9 THE SKILL THIS TEACHER HAS TO GET STUDENTS TO THINK FOR THEMSELVES?
(Are students' ideas and opinions worth something in this class? Do students help decide how to solve problems and how to get their work done? Do they get at the real reasons why certain things happen?)
Below Average   Average   Good   Very Good   The Very Best

10. THE GENERAL (ALL-ROUND) TEACHING ABILITY OF THIS TEACHER?
(All factors considered, how close does this teacher come to your ideal?)
Below Average   Average   Good   Very Good   The Very Best

(over)

11. PLEASE NAME ONE OR TWO THINGS THAT YOU ESPECIALLY LIKE ABOUT THIS TEACHER.

12. PLEASE GIVE ONE OR TWO SUGGESTIONS FOR THE IMPROVEMENT OF THIS TEACHER.

13. PLEASE NAME ONE OR TWO THINGS THAT YOU ESPECIALLY LIKE ABOUT THIS COURSE.

14. PLEASE GIVE ONE OR TWO SUGGESTIONS FOR THE IMPROVEMENT OF THIS COURSE.

Prepared by the Student Reaction Center, Division of Field Services, Western **Michigan University, Kalamazoo, Michigan.**

APPENDIX C

SUMMARY OF COMMENTS BY STUDENTS

Teacher "Y" Class Drafting & Design II Date 2/18/65

1. Things students especially like about this teacher.

He knows his subject matter well . . . Is patient and understanding . . . Willingness to give extra help when needed . . .Fairness . . .Makes you think . . .Considerate . .

2. Suggestions for the improvement of this teacher.

His ability to express himself better . . . Try to maintain a more definite pattern rather than skipping around . . . Should try to develop a better method of presentation . . . More of an open mind concerning students contributions and ideas . . . Better planned work schedule.

3. Things students especially like about this course.

The course is interesting . . . The course covers a large variety of subject matter . . . Structural drafting . . . Using board skills with design skills . . .

4. Suggestions for the improvement of this course.

More visual aids (Motion pictures) . . . More time to finish class projects . . . More field trips to industry. Course outline jumps around too much . . . More illustrations and visuals.

NOTE: The three dots separate the comment by one student from that by another.

APPENDIX D

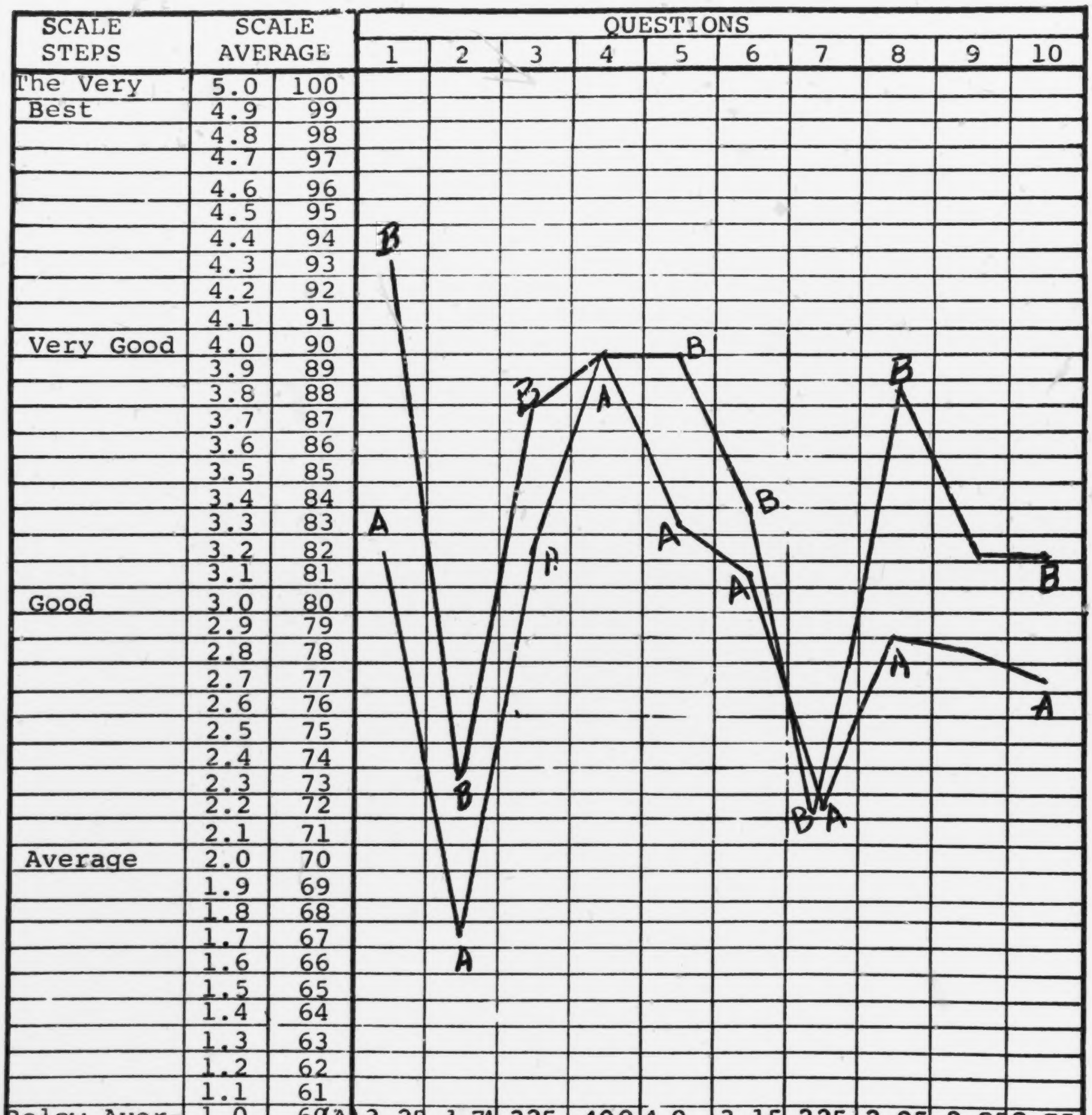
STUDENT REACTION REPORT

Teacher "Y" No. ____ Date ________

Class	Hour	Subject
A	AM	Drafting & Design II
B	PM	Drafting & Design I
C		

KEY TO QUESTIONS

1. Knowledge
2. Explanations
3. Fairness
4. Discipline
5. Sympathy
6. Amount of Learning
7. Interesting
8. Businesslike
9. Think
10. All-round ability



	1	2	3	4	5	6	7	8	9	10
(A)	3.25	1.74	3.25	4.00	4.0	3.15	2.25	2.95	2.85	2.75
(B)	4.34	2.35	3.72	4.00	3.3	3.40	2.21	4.02	3.21	3.22

NUMBER OF STUDENTS WHO CHOSE EACH SCALE STEP

SCALE STEPS	CLASS A										CLASS B										CLASS C									
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
The V.Best				1	1			1			7		4	3	5	3		3	1	1										
Very Good	5		5	6	7	5	1	3	4	2	8	4	6	11	3	5	7	11	9	7										
Good	5	1	5	5	4	4	3	4	4	5	2	3	3	3	6	4	3	2	2	7										
Average	2	7	2	1		2	6	2	2	5		5	4		3	4	4	1	3	2										
Below Ave.		4				1	2	2	2			5				1	3		1											
TOTALS	12	12	12	12	12	12	12	12	12	12	17	17	17	17	17	17	17	17	17	17										

APPENDIX E

PERSONAL DATA

( ) Pretest _____

( ) Posttest _____

( ) Number - Name ________________________

( ) Subject taught ________________________

( ) Years of teaching experience _________

( ) Age _____

( ) Assigned group ________________________

( ) School ________________________________

( ) Number of students _____

NOTE: Do not fill in parenthesis, use line on right.

APPENDIX F

STATE OF NEW JERSEY
DEPARTMENT OF EDUCATION
VOCATIONAL DIVISION
TRENTON 08625

TO:

FROM: Wilmot F. Oliver

DATE:

RE: Student and/or Supervisory Informational Feedback

The enclosed informational feedback is for your information only. No one besides the researcher will ever see this information without your expressed consent. The object of this research is to give you an opportunity to see yourself as your students, or your supervisor, sees you. It is hoped that you will carefully consider this information and then act in accordance with your best professional judgement. Only you can judge how valid this information is, and only you can benefit from utilizing it to improve your teaching image.

In approximately 36 weeks the student opinion questionnaire will be administered for the third time. We are anxious to determine whether there is any significant change in the second and third administration as a result of putting this information at your disposal.

You will receive the student and/or supervisor reaction reports as soon as they are processed. You will also receive a summary of the findings of this research project when all the data has been analyzed and the final report written. We trust that this information will be helpful in your endeavor to increase your teaching effectiveness.

Thanks for your cooperation.

APPENDIX G

INSTRUCTIONS FOR ADMINISTERING THE STUDENT OPINION QUESTIONNAIRE

1. It is very important that someone other than the teacher receiving the service should be in charge of the class. That "someone" could be the shop teacher next door or any member of the staff.

2. After the student-opinion questionnaire is handed out to the students, they should be instructed to carefully read the directions at the top of the sheet. Emphasize that they are not to write their names on the paper, and that they will remain completely anonymous.

3. The students should be told that this questionnaire is part of a research project and is designed to help their instructor, and that the summary of the information contained in questionnaires will be released to no one except the instructor.

4. These opinion questionnaires will be administered to twenty to thirty students for each shop teacher. Grades 10,11, and 13 will be used for all teachers who have these grades. In the event a teacher has only grades 9 or 12, these will be used in the study. Only shop teachers will be used in this research.

5. Immediately upon collecting the questionnaires from the students, they will be sealed in the envelope provided for the purpose and forwarded to Mr. Oliver at Rutgers University.

6. Upon receipt of the answered questionnaires at Rutgers, they will be converted into a report and mailed to the teacher along with interpretive information.

7. Be sure to fill in the teacher's name, subject and grade level, years of teaching experience, and his age upon the personal data form provided with each envelope.

8. All unused student-opinion questionnaires should be returned in the envelope.

9. Have students indicate their grade level in the upper right-hand corner of the questionnaire.

APPENDIX H

STATE OF NEW JERSEY
DEPARTMENT OF EDUCATION
VOCATIONAL DIVISION
TRENTON 08625

TO:

FROM: Wilmot F. Oliver

DATE:

Will you kindly have your students fill out the enclosed Student-Opinion Questionnaire on or about
This is for the post-test which will be the final phase of this research project.

You will receive a summary of these reactions as soon as the questionnaires are processed by our computer center at Rutgers.

Thanks again for your cooperation in this project.

APPENDIX I

STATE OF NEW JERSEY
DEPARTMENT OF EUDCATION
VOCATIONAL DIVISION
TRENTON 08625

TO:

FROM: Wilmot F. Oliver

DATE:

Will you kindly fill out the enclosed Supervisor-Opinion Questionnaires. This is for the post-test which will be the final phase of this research project.

A summary of these reactions will be sent to the teachers in experimental groups three (3) and five (5).

Thanks again for your cooperation.

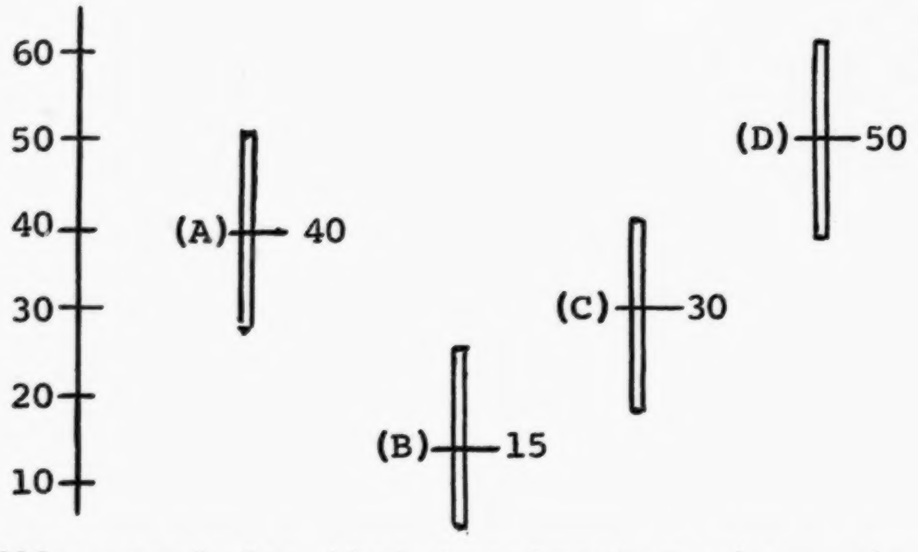
APPENDIX J

CONFIDENCE INTERVALS

H.P. Andrews

Frequently experimenters, or statisticians, find it convenient or desirable to summarize the results of an experiment by tabulating the experimental treatment means along with their 95% Confidence Intervals. Sometimes it is desirable to present these graphically and to plot the means and confidence intervals as bar graphs so as to gain some insight into the relationships among the treatment means. In either case, for purposes of interpretation, one scans the confidence intervals and readily concludes that those intervals which do not overlap one another reveal that these treatments differ significantly from one another.

For example, if one scans the following array of confidence intervals



One readily concludes that treatment B gives significantly lower results than treatments A and D, that treatment C gives significantly lower results than treatment D--but what about C with respect to A, or with respect to B; and

A with respect to D? Are those differences statistically significant? Should the fact that their confidence intervals overlap be properly interpreted that the differences between those means are not statistically significant? Not necessarily--it turns out that two 95% confidence intervals may overlap to some extent and the difference between the means still be significant at P=5%.

The question of significance (statistical) can, of course, be answered by specific tests of significance, e.g. the student's "t" test; and some experimenters like to use the "statistical Yardsticks": $LSD = 5(s_{\bar{d}})$, or when multiple comparisons are of interest some multiple range test or the $JSD = q(s_{\bar{x}})$ as "go or no go" gauges of statistical significance.

If we wish to be led to the same conclusions using confidence intervals, either graphically or in tabular form, we need to construct confidence intervals which when they overlap at all tell us that the treatments do not differ significantly. Then the criterion of no overlapping provides an "exact" method of ascribing statistical significance to differences between treatment means at the prescribed probability level. When in effect what is needed is a Least Significant Confidence Interval (LSCI consistent with LSD) or a Just Significant Confidence Interval (JSCI consistent with JSD).

If two means differ by an amount just equal to or greater than the LSD or JSD are considered significantly different (at the selected alpha level), then an interval of one half the LSD or JSD when constructed about the means would show no overlapping. This failure to overlap then would lead to the same conclusion that the means differed significantly. Similarly, two means bounded by this interval would show overlapping which would lead to the correct conclusion that those means were not significantly different, at the prescribed level or probability.

Accordingly:

$$LSD_g = t_g s_{\bar{d}} = t_g \sqrt{\frac{2s^2}{N}} = t_g \sqrt{2}\; s/\sqrt{N} = \underline{\underline{\sqrt{2}\, t_g s_{\bar{x}}}}$$

$$LSCI_g = LSD_g/2 = \pm\sqrt{2}/2\; t_g s_{\bar{x}} = \underline{\underline{\pm\, 0.707\, t_g s_{\bar{x}}}}$$

OR

$$JSD_g = \underline{\underline{\frac{q}{x}\, g^{K} s_{\bar{x}}}}$$

$$JSCI_g = \underline{\underline{\frac{q}{x}\, g^{K} s_{\bar{x}}/2}}$$

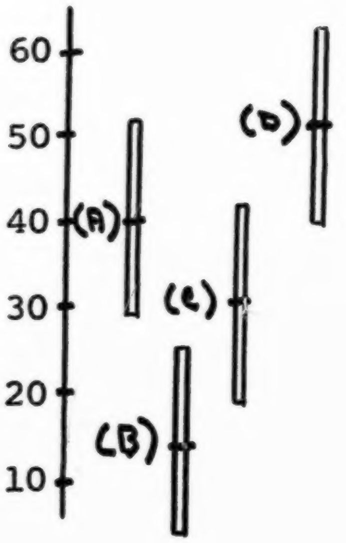
In the example shown on page 1, consider:

$s_{\bar{x}} = 5$, $t_{.05} = 2.00$

then $LSD_{.05} = 2\ (2)\ (5) = 14.14$

$LSCI_{.05} = 0.707\ (2)\ (5) = \pm\ 7.07$

and these LSCI shown graphically reveal:



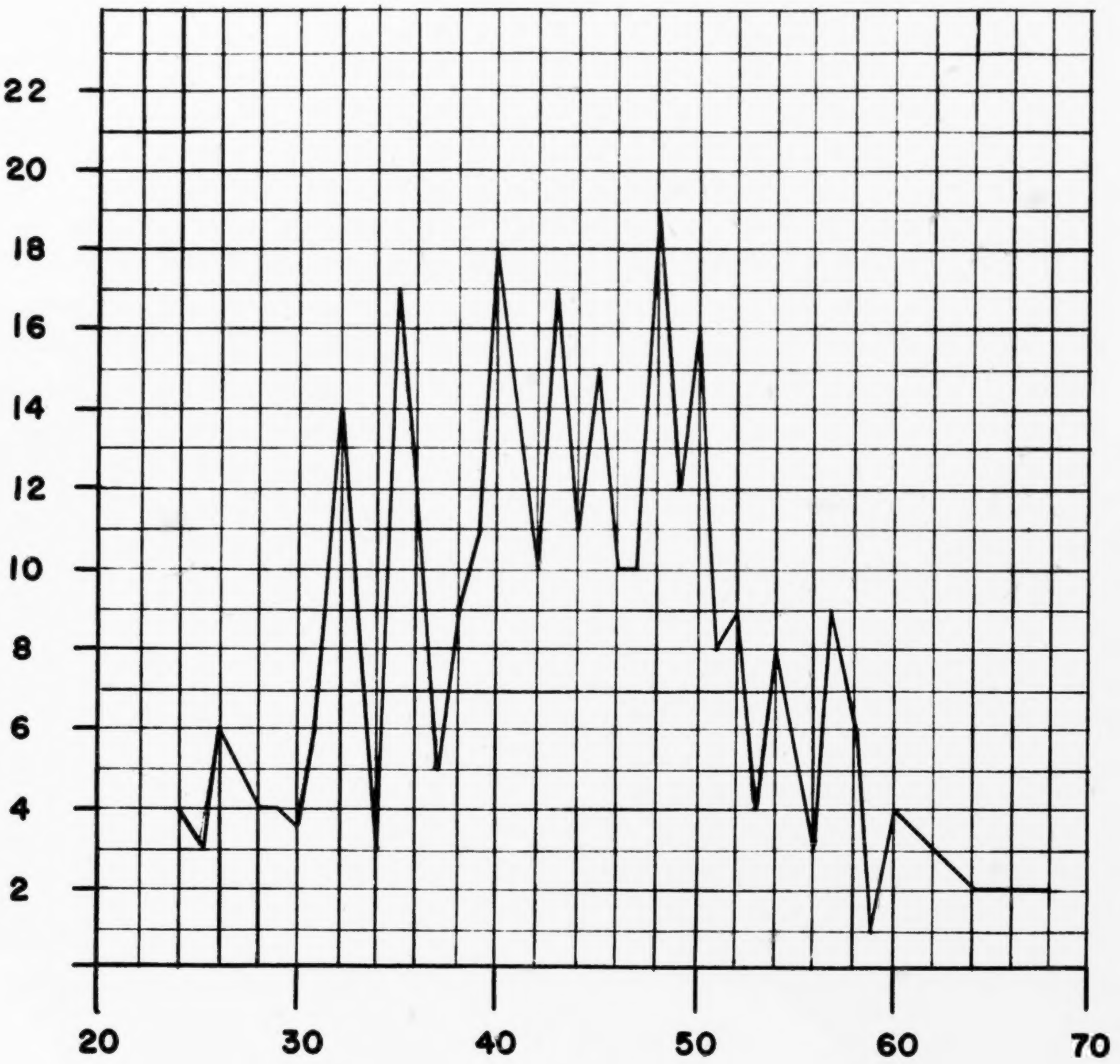
We will still conclude that Treatment B gives significantly lower results than A and D, and that Treatment C gives significantly lower results than D. But now we find that Treatment B also gives significantly lower results than C--and conclude that Treatments A and C are not significantly different and that Treatments A and D are not significantly different. The Least Significant Confidence Intervals therefore provide us with a specific graphical method of stating the statistical significance of the results while giving, at the same time, a picture of the general relationships among the treatment means.

APPENDIX K

SCHOOLS PARTICIPATING IN THE STUDY

Atlantic City
Bayonne
Belvidere
Bergen County Vocational School
Bloomfield High School
Bridgeton
Buck High School
Bucks County Vocational School
Burlington County Vocational School
Camden County Vocational School
Cape May County Vocational School
Deptford High School
Essex County Vocational School
Flemington
Fletcher Brown Vocational School
Freehold Regional
Glassboro
Middlesex County Vocational School
Monmouth County Vocational School
Morris County Vocational School
New Brunswick High School
North Hunterdon Regional
Northern Valley Regional
Nutley
Ocean County Vocational School
Passaic County Vocational School
Rockland County Vocational School
Somerset County Vocational School
Thomas Edison Vocational School
Trenton Central High School
Union City
Union County Regional
Union County Vocational School
Warren County Vocational School
Washington High School

APPENDIX L



GRAPH I

AGE OF INSTRUCTORS

DISTRIBUTION OF AGES IN CONTROL AND EXPERIMENTAL GROUPS.

APPENDIX M

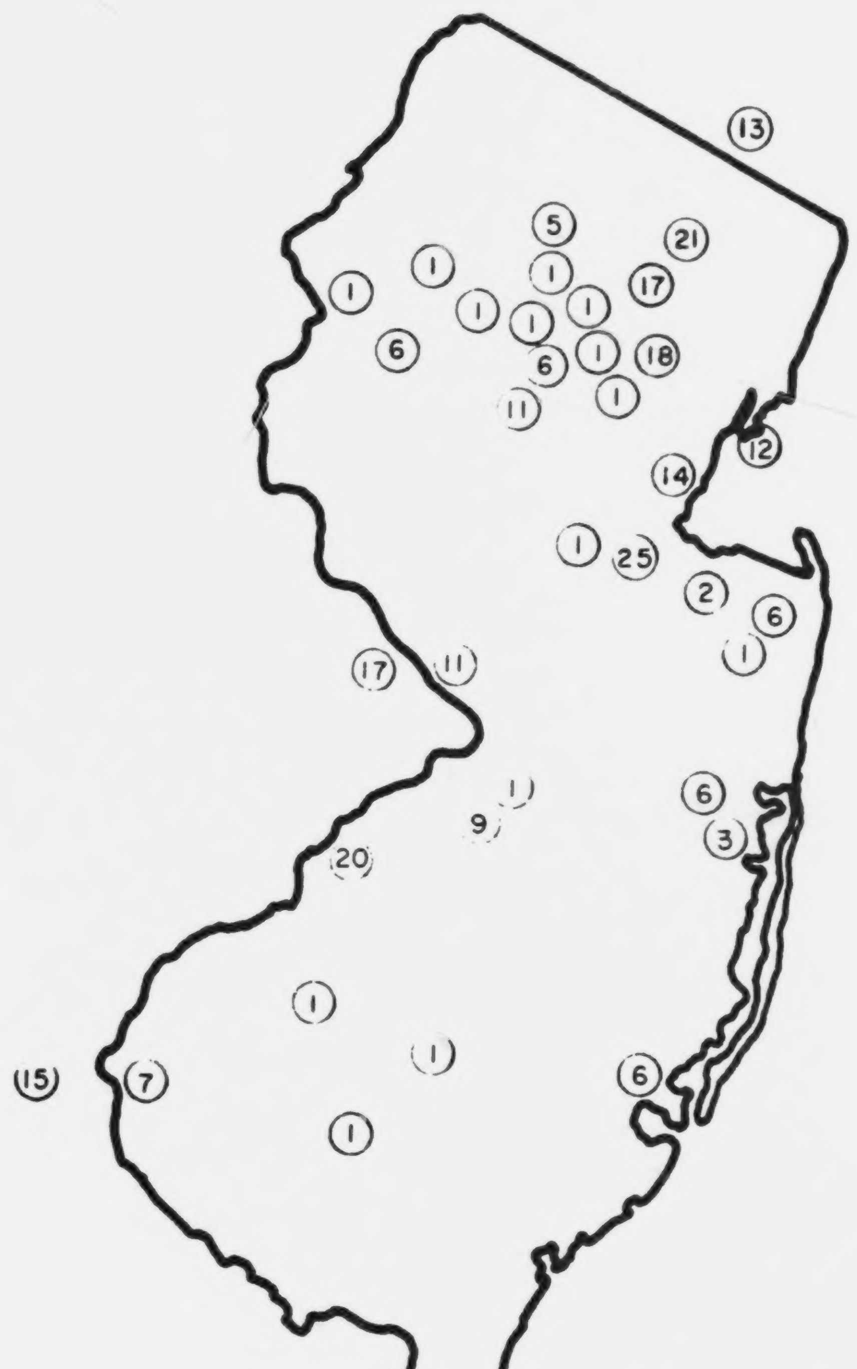

13
5
21
1
1
17
1
1
1
1
6
6
1
18
1
11
12
14
1
25
2
6
1
17
11
1
6
9
3
20
1
1
15
7
6
1
8

FIGURE 2

LOCATIONS OF SCHOOLS SUPPLYING VOCATIONAL TEACHERS PARTICIPATING IN THE STUDY